Copyright
by
Bowen Hua
2018

The Dissertation Committee for Bowen Hua
certifies that this is the approved version of the following dissertation:

Unit Commitment: Tight Formulation and Pricing

Committee:

Ross Baldick, Supervisor

Jonathan Bard

Constantine Caramanis

Surya Santoso

Hao Zhu

Unit Commitment: Tight Formulation and Pricing

by

Bowen Hua

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2018

Acknowledgments

I would like to first express my gratitude to my advisor, Dr. Ross Baldick, for his guidance and support. Dr. Baldick is an inspiring mentor, an engaging teacher, and a true intellectual. I am extremely fortunate to have had the opportunity to work with him. I cannot imagine a better advisor.

My gratitude extends to my committee members. I would like to thank Dr. Jonathan Bard for introducing me to the field of integer programming and sharing his insights on bilevel optimization. I am grateful to Dr. Surya Santoso for offering the power quality class and for his helpful comments on my research. I thank Dr. Hao Zhu and Dr. Constantine Caramanis for their constructive advice on my dissertation.

I am thankful to professors and colleagues outside of the University of Texas with whom I collaborated. Dr. Yonghong Chen and Dr. Congcong Wang from Midcontinent ISO provided guidance on the practical side of power system economics and shared valuable insights on the topics of pricing and combined-cycle modeling. I am grateful to Dr. Dane Schiro, Dr. Tongxin Zheng, and Dr. Eugene Litvinov from ISO New England for our collaboration on the pricing of multi-interval real-time markets. Dr. Kevin Wood from Naval Postgraduate School offered inspiration and enlightening comments on interdiction problems. I thank Dr. Jianhui Wang from Argonne National Lab-

oratory for helpful advice on generation expansion planning.

I am also indebted to my colleagues in and out of our research group for benefiting from our interactions. I am especially grateful to Mohammad Majidi, Yezhou Wang, Bing Huang, Manuel Garcia, Xinbo Geng, Tong Zhang, Piyapath Siratarnsophon, Juan Andrade, and Sambuddha Chakrabarti for discussions and ideas.

Finally, this dissertation would not have been completed without the love and support from my parents. I dedicate this dissertation to them.

Unit Commitment: Tight Formulation and Pricing

Publication No. _____

Bowen Hua, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Ross Baldick

In recent years, with increasing renewable supply variability, thermal power plants have started up and shut down more frequently. These discrete commitment decisions, optimized in the unit commitment (UC) problem, have an impact on system operations as well as generation expansion planning (GEP). The non-convex costs associated with the commitment decisions may also lead to generators' incentive to deviate from the optimal dispatch under locational marginal prices. In this dissertation, we first propose a convex relaxation of UC based on a primal formulation of the Lagrangian dual problem. This convex relaxation is used (i) to solve the convex hull pricing problem in polynomial time, providing prices with better incentives in non-convex electricity markets, and (ii) to construct a computationally efficient GEP model that represents operational flexibility limits. Next, we present a tight formulation for the commitment of combined-cycle units with representation of their transition ramping. Finally, we propose a pricing method that reduces out-of-market payments in multi-interval real-time markets.

Table of Contents

List of Tables

List of Figures

Part I

Introduction

Chapter 1

Introduction

1.1 Unit Commitment and Non-Convex Electricity Markets

With high shares of variable renewable energy sources and increasing supply variability, thermal power plants in electric power systems have seen more frequent start-ups and shut-downs [73]. These start-up and shut-down decisions, also known as commitment decisions, are solved in the unit commitment (UC) problem, an optimization problem that schedules a portfolio of generators to satisfy the demand for electricity over a certain number of periods. Because of the indivisibility of generators and their non-negligible non-convex costs,[1] the UC problem is typically formulated as a mixed-integer program. Because thermal generators may take hours to be brought online, UC is solved day-ahead or hour-ahead to deploy enough system resources to adequately address the real-time uncertainty of demand and renewable generation.

In addition to making UC a computationally difficult mixed-integer

[1] As examples: start-up costs, the costs to bring the boiler, turbine, and generator from shut-down conditions to a state ready to connect and be synchronized to the system; and, no-load costs, the costs to maintain the boiler operating and the turbine and generator spinning at synchronous speed.

program, the indivisibility of generators has economic implications. Day-ahead and some real-time electricity markets in the US currently base their market clearing model on a UC problem [22,67,78]. The independent system operator (ISO) who coordinates, controls and monitors the operation of the electrical power system as well as the electricity market sends energy prices and target quantity instructions to each generating unit (called "unit" for short hereafter) based on a welfare-maximizing solution to the UC problem. From a micro-economic viewpoint, the ISO's UC problem is a social planner's problem.

Ideally, energy prices provide incentives for profit-maximizing market participants to comply with the ISO's commitment and dispatch decisions. Various issues prevent this ideal, however, including the indivisibility, or more generally, non-convexities of the underlying UC problem that arise from units' operating characteristics. Consequently, start-up and no-load costs of units may not be covered by sales of energy at locational marginal prices (LMPs). More generally, in a market with non-convexities, there might be no set of uniform prices[2] that supports a welfare-maximizing solution [29].[3]

One way to address this problem is to maintain uniform energy prices based on marginal energy costs and provide side payments to units that have an incentive to deviate from the ISO's solution. This side payment is also known as an "uplift" payment. In principle, the amount of uplift payment to

[2]A uniform price is a single price that applies to all transactions at a given bus. Prices may vary locationally.

[3]A set of prices is said to support a solution if the economic agents' profit-maximizing decisions align with this solution.

a unit should cover its lost opportunity cost, the gap between its maximum possible profit and the actual profit obtained by following the ISO's solution. For example, the cost of a fast-start unit dispatched at its minimum limit may not be covered by sales of energy at its marginal cost. If revenues are based solely on LMP, then the profit-maximizing decision of this unit is to shut down. An uplift payment is needed to keep this unit online [82]. Unlike energy prices, uplift payments are non-uniform (discriminatory) in that the amount of payment is unit-specific. These side payments make it harder for a potential entrant to determine if new entry would be profitable, particularly if the uplift payments are not disclosed publicly.

Transparency of the market can be improved by keeping uplift payments as low as possible. To this end, convex hull pricing [29, 34, 70] has been proposed as a pricing scheme that minimizes certain uplift payments over all possible uniform prices, and has received much attention. The convex hull prices are the dual maximizers of the Lagrangian dual problem of UC. Determining these dual maximizers is computationally expensive [33, 82]. In Chapter 2 and Chapter 3 of this dissertation, we address this computational issue by proposing a polynomially-solvable primal formulation for convex hull pricing.

1.2 Unit Commitment and Generation Expansion Planning

Generation expansion planning (GEP) chooses investments in new generating units to meet load growth and replace retirements. Vertically integrated utilities use GEP to minimize investment and operational costs in a centralized manner. In the context of a deregulated electricity sector, GEP gives an optimal generation mix, providing the ideal outcome of a capacity market and informs regulatory decisions such as renewable portfolio standards.

High penetration of renewable generation changes the short-term scheduling of thermal power plants by increasing the requirement for operational flexibility—the ability to cope with variability and uncertainty in both generation and demand [48]. Recent studies [41, 60, 76] show that operational flexibility has a considerable impact on GEP. Ignoring limits on operational flexibility (e.g., indivisibility of the generators, minimum run time constraints) in GEP can lead to a suboptimal expansion plan and higher operational costs.

To fully represent limits on operational flexibility, the embedded operational model needs to be a UC problem that represents chronological demand and renewable generation. However, the resulting GEP model is computationally challenging. The Lagrangian dual problem of UC provides a tight relaxation, as the relative duality gap between the UC problem and its Lagrangian dual approaches zero as the number of heterogeneous generators approaches infinity [7, 8]. In Chapter 4 of this dissertation, following the line of research that simplifies the operational UC model, we present a GEP model

that adopts the primal formulation of the Lagrangian dual of UC as the embedded operational model. Our model is computationally efficient, because the embedded operational model is a continuous and polynomially-solvable optimization problem; the only integer variables in our GEP model are the investment decisions.

1.3 Commitment and Transition of Combined-Cycle Units

A combined cycle unit (CCU) is a type of generator that consists of one or more combustion turbines (CTs), each with a heat recovery steam generator (HRSG), as well as one or more steam turbines (STs). Based on different combinations of CTs and STs, a CCU can be operated in one of several configurations. This is different from simple-cycle units for which we assume binary on/off states.

Because of the increasing penetration of renewables, many combined cycle units have been operated as providers of flexibility in power systems. Whereas in the past CCUs might run for lengthy periods of time serving base load, they now respond quickly to variations in renewable supply and demand by frequent ramping, as well as more frequent transitions between configurations. In addition to determining the configuration of CCUs in day-ahead unit commitment, optimizing the configuration of CCUs in the look-ahead commitment and dispatch can help meet projected system conditions by pre-positioning CCUs to cope with the variability of renewables.

Most existing models for CCUs assume that CCUs start/end their pro-

duction within one interval. This assumption can often be violated, especially in look-ahead commitment and dispatch models where the length of the interval is typically fifteen minutes. To accurately model the transitions, Chapter 5 proposes a mixed-integer programming model for CCUs where we explicitly represent their transition ramping.

1.4 Multi-Interval Real-Time Markets

Several wholesale markets in the US have implemented a *multi-interval real-time market* (MIRTM) to cope with the variabilities of renewable resources. Instead of solving a single-interval economic dispatch problem, MIRTM is based on a *look-ahead dispatch* (LAD) problem that considers several intervals. By expanding the time horizon, MIRTM allows for more efficient dispatch of generation to meet projected system conditions by pre-positioning resources to cope, for example, with large ramps in net load [32, 88, 89].

In a typical implementation of MIRTM, the ISO solves LAD in a fashion that resembles model predictive control (MPC) [89]. In the MPC-like implementation, the collection of settlement prices from a sequence of LAD problems does not necessarily support the ISO's dispatch. As a result, a generator might have an incentive to deviate from the ISO's dispatch. Fundamentally, this is the result of each successive optimization problem treating historical losses as sunk costs.

To solve the incentive issues, we present a new multi-interval pricing method in Chapter 6. We show the theoretical incentive properties of the

proposed method with a fixed finite horizon and perfect foresight. We also perform numerical analysis on an ISO New England-based system.

1.5 Outline

Chapter 2 examines the Lagrangian dual problem of UC from a theoretical perspective. We first introduce the UC problem and its Lagrangian dual. We propose a polynomially-solvable primal formulation for the Lagrangian dual problem. This formulation explicitly describes, for each generating unit, the convex hull of its feasible set and the convex envelope of its cost function.

Chapter 3 applies our primal formulation to convex hull pricing. We first introduce the convex hull pricing problem in non-convex electricity markets. Applying our primal formulation leads to exact convex hull prices absent ramping constraints, and a tractable approximation when ramping constraints are present. Using several case studies, we demonstrate the computational efficiency of our method. We show that our method leads to reduced uplift payment compared to an approximation of convex hull pricing used by Midcontinent ISO (MISO).

Chapter 4 applies our primal formulation to GEP. We present a GEP model that adopts our primal formulation as the embedded operational model. Our model is computationally efficient, because the embedded operational model is a continuous and polynomially-solvable optimization problem; the only integer variables in our GEP model are the investment decisions. We demonstrate the performance of our GEP model and the tightness of our

convex relaxation using a Texas system in which we consider a chronological load curve of 8760 hours.

Chapter 5 presents a tight formulation for the combined-cycle units with representation of their transition ramping. We first identify the drawbacks of the existing methods that assume the completion of any transition within a single interval. We then remove this invalid assumption and propose a formulation with explicit representation of transition ramping. We show theoretical results on the tightness and compactness of our formulation.

Chapter 6 presents an existing proposal and a new multi-interval pricing method for MIRTM. We show that both methods incorporate historical losses and produce prices that appropriately incentivize generators to follow multi-interval dispatch instructions when implemented with a fixed finite horizon and perfect foresight. Realistic rolling horizon implementations for an ISO New England-based system suggest that the new method results, in practice, in lower out-of-market payments than both current ISO implementations and the existing proposal.

We conclude this dissertation in Chapter 7 with a summary of results and future research directions.

Part II

A Convex Relaxation of Unit Commitment and its Applications

Chapter 2

A Convex Relaxation of Unit Commitment

2.1 Introduction

In electric power system operations, the unit commitment (UC) problem schedules a portfolio of generators to supply the demand for electricity over a certain number of periods at maximum social welfare. The constraints include technical limits of each generators (called *private constraints* hereafter), as well as *system-wide constraints* that are complicating, such as power balance constraints and transmission constraints.

In the context of a vertically-integrated utility, the monopolistic operator solves the UC problem for an optimal operational schedule. In the context of a deregulated electricity sector, an independent system operator (ISO) has to determine the set of accepted supply offers and demand bids submitted from the market participants, as well as the energy prices to be used for settlements. To this end, the ISO solves an offer-based UC model that can be viewed as the social planner's model in microeconomics.

Because of the indivisibility of the generating units, UC is a mixed-

This chapter is based in part on the following publication to which the coauthors contributed equally: Bowen Hua and Ross Baldick. "A convex primal formulation for convex hull pricing." *IEEE Transactions on Power Systems* 32.5 (2017): 3814-3823.

integer and non-convex problem that is difficult to solve, especially for large-scale problems. Since UC is separable across generators absent the system-wide constraints, the Lagrangian relaxation method has in the past been used to solve UC [5, 55]. Since the early 2000s, because of significant improvements in MIP solvers, mixed-integer programming (MIP) became the dominant approach to solve UC problems [59].

Despite no longer being the dominant approach to solving UC, Lagrangian relaxation of UC remains an interesting problem because of two reasons: the Lagrangian dual problem produces dual maximizers that are efficient prices in a nonconvex electricity market; and the Lagrangian dual problem itself is a tight convex relaxation of UC.

In this chapter, we first describe a polyhedral study. The results of this study will be utilized in Chapter 3 when characterizing the primal formulation of the Lagrangian dual, and in Chapter 5 when we study the feasible region of a combined cycle unit. We then introduce the UC problem and its Lagrangian dual in which we dualize the system-wide constraints. Finally, we propose a polynomially-solvable primal formulation for the Lagrangian dual problem. Part of this work is also reported in [37].

2.2 Polyhedral Study of Some Mixed-Integer Sets

In this section, we develop a polyhedral study of two types of mixed-integer sets. We present two new theorems that will be utilized when we characterize the convex hull of the feasible sets of each individual generator.

The convex hull of a compact set defined by linear inequalities and integrality requirements is a bounded polyhedron. In general, it is difficult to obtain an explicit description of the convex hull of a mixed-integer set defined by arbitrary linear constraints. The number of valid inequalities needed is typically exponential in the size of the input [16]. In the following, we will describe special cases of mixed-integer sets for which it is easy to find their convex hulls. These cases have specific relevance to the unit commitment problem, but are also more widely applicable. We therefore develop the cases in a general format.

2.2.1 Polyhedral Cone with an Integer Scalar

Let $\boldsymbol{y} \in \mathbb{R}^{n_y}$ be a continuous vector and let $z \in \mathbb{Z}_+$ be a nonnegative integer scalar. Let $\mathcal{Z} \subset \mathbb{Z}_+$ be a finite set of nonnegative integer scalars. Suppose we have the following mixed-integer set:

$$\mathcal{F}_1 = \{(\boldsymbol{y} \in \mathbb{R}^{n_y}, z \in \mathcal{Z}) \,|\, A\boldsymbol{y} \leq z\boldsymbol{b}\}, \tag{2.1}$$

where $A \in \mathbb{R}^{n_a \times n_y}$ and $\boldsymbol{b} \in \mathbb{R}^{n_y}$. Without the integrality requirement, this set would have been a polyhedral cone.[1]

Let $\text{conv}(\cdot)$ denote the convex hull of a set.[2] Let $\min \mathcal{Z}$ and $\max \mathcal{Z}$, respectively, be the minimum and maximum elements in the set $\mathcal{Z}$. We show that the convex hull of $\mathcal{F}_1$ can be easily characterized.

[1] A set $\mathcal{C}$ is called a cone if for every $x \in \mathcal{C}$ and $\theta \geq 0$ we have $\theta x \in \mathcal{C}$. A cone $\mathcal{C}$ is called a polyhedral if there is a matrix A such that $\mathcal{C} = \{x \,|\, Ax \geq 0\}$. For the cases in this chapter, $x = (y, z)$.

[2] The convex hull of a set is all convex combinations of points in that set.

Theorem 2.1. *The convex hull of $\mathcal{F}_1$ is*

$$\mathcal{G}_1 = \{(\boldsymbol{y} \in \mathbb{R}^{n_y}, z \in \mathbb{R}) \mid A\boldsymbol{y} \leq z\boldsymbol{b}, \min \mathcal{Z} \leq z \leq \max \mathcal{Z}\}.$$

Proof. Because

$$\mathcal{F}_1 = \{(\boldsymbol{y} \in \mathbb{R}^{n_y}, z \in \mathbb{Z}) \mid A\boldsymbol{y} \leq z\boldsymbol{b}, \min \mathcal{Z} \leq z \leq \max \mathcal{Z}\},$$

it suffices to show that z^* is integral for all extreme points $(\boldsymbol{y}^*, z^*)$ of $\mathcal{G}_1$ [14, Theorem 4.3]. We prove this by contradiction.

Suppose not. Then there exists an extreme point of $\mathcal{G}_1$, $(\boldsymbol{y}^*, z^*)$, with fractional z^*. Since an extreme point lies within $\mathcal{G}_1$, we have $A\boldsymbol{y}^* \leq z^*\boldsymbol{b}$ and $\min \mathcal{Z} < z^* < \max \mathcal{Z}$, since $\min \mathcal{Z}$ and $\max \mathcal{Z}$ are integer-valued. The following inequalities hold for any $1 > \epsilon_1 > 0$ and $1 > \epsilon_2 > 0$:

$$\begin{aligned} A(1+\epsilon_1)\boldsymbol{y}^* &\leq (1+\epsilon_1)z^*\boldsymbol{b}, \\ A(1-\epsilon_2)\boldsymbol{y}^* &\leq (1-\epsilon_2)z^*\boldsymbol{b}. \end{aligned}$$

Now pick $1 > \epsilon_1 > 0$ and $1 > \epsilon_2 > 0$ such that the following inequalities hold:

$$\min \mathcal{Z} \leq (1+\epsilon_1)z^* \leq \max \mathcal{Z}, \tag{2.2}$$
$$\min \mathcal{Z} \leq (1-\epsilon_2)z^* \leq \max \mathcal{Z}. \tag{2.3}$$

We claim that such ϵ_1 and ϵ_2 always exist. For example, $\epsilon_1 = \min\{\frac{\max \mathcal{Z}}{z^*} - 1, 0.5\}$ and $\epsilon_2 = \min\{1 - \frac{\min \mathcal{Z}}{z^*}, 0.5\}$ would suffice. This is because (i) the left-hand inequality of (2.2) and the right-hand inequality of (2.3) are satisfied for

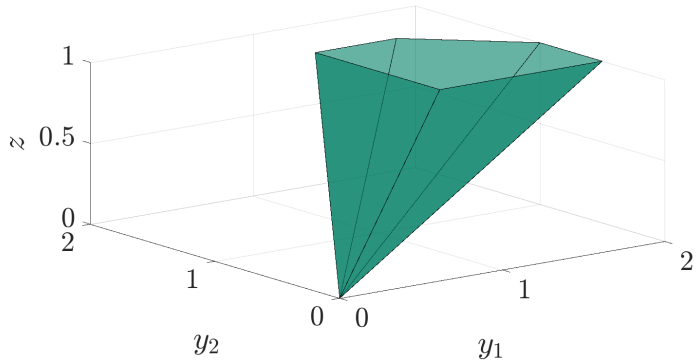


Figure 2.1: An example for which Theorem 2.1 applies. The convex hull description of this mixed-integer set is trivial.

any $\epsilon_1 > 0$ and $\epsilon_2 > 0$, respectively, and (ii) the right-hand inequality of (2.2) and the left-hand inequality of (2.3) are satisfied by construction of ϵ_1 and ϵ_2, respectively.

The above inequalities imply that both $(1+\epsilon_1)(\boldsymbol{y}^*, z^*)$ and $(1-\epsilon_2)(\boldsymbol{y}^*, z^*)$ are in $\mathcal{G}_1$. As a result, $(\boldsymbol{y}^*, z^*)$ is a convex combination of two points in $\mathcal{G}_1$, namely $(1+\epsilon_1)(\boldsymbol{y}^*, z^*)$ and $(1-\epsilon_2)(\boldsymbol{y}^*, z^*)$. This contradicts $(\boldsymbol{y}^*, z^*)$ being an extreme point. □

Fig. 2.1 shows an example for which Theorem 2.1 applies. In this example, $n_y = 2$ and $\mathcal{Z} = \{0, 1\}$. The mixed-integer set is expressed as:

$$\begin{aligned}\{(\boldsymbol{y} \in \mathbb{R}^2, z \in \{0,1\}) \,|\, & z \le y_1 \le 2z, \\ & 0.5z \le y_2 \le 1.5z, \\ & 0 \le y_1 + y_2 \le 3z\}.\end{aligned}$$

Its convex hull is simply:

$$\{(\boldsymbol{y} \in \mathbb{R}^2, z \in \mathbb{R}) \,|\, z \leq y_1 \leq 2z,$$
$$0.5z \leq y_2 \leq 1.5z,$$
$$0 \leq y_1 + y_2 \leq 3z,$$
$$0 \leq z \leq 1\}.$$

In the context of unit commitment, the case considered in Theorem 2.1 is sufficiently general to represent the feasible region of a single-cycle generator in a single period. The integer variable z represents the commitment status of the generator, and the continuous vector represents the power output and ancillary services provision of the generator.

2.2.2 Polyhedral Cone with an Integer Vector

Consider the following polyhedral cone defined by double-sided inequalities:

$$\mathcal{C} = \{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \,|\, B\boldsymbol{z} \leq A\boldsymbol{y} \leq D\boldsymbol{z}\}, \tag{2.4}$$

where $A \in \mathbb{R}^{n_a \times n_y}$, where $B, D \in \mathbb{R}^{n_a \times n_z}$, and where $\boldsymbol{z}$ is redefined to be a vector. Without loss of generality, assume that $\mathcal{C}$ is full-dimensional. If not, we can make the polyhedron full-dimensional by eliminating some variables. We can also assume without loss of generality that coefficient matrix of the inequalities defining $\mathcal{C}$ has linearly independent rows.[3] If not, since we do not have constant terms in the inequalities that define the polyhedral cone $\mathcal{C}$,

[3]That is, we require that $\begin{bmatrix} -A & B \\ A & -D \end{bmatrix}$ has linearly independent rows.

either we have at least one redundant constraint that we can remove, or we have an equality that violates the full-dimensionality assumption.

We consider the following set of (finitely many) feasible $\boldsymbol{z}$ vectors in the non-negative orthant:

$$\mathcal{P} \subseteq \mathbb{Z}_+^{n_z}.$$

Suppose we have the following polyhedron:

$$\mathcal{P}_r = \{\boldsymbol{z} \in \mathbb{R}^{n_z} \mid H\boldsymbol{z} \leq \boldsymbol{h}, \boldsymbol{z} \geq \boldsymbol{0}\}. \tag{2.5}$$

If $\text{conv}(\mathcal{P}) = \mathcal{P}_r$, and if $\mathcal{C}$ satisfies a certain condition, we show that the convex hull of $\mathcal{C} \cap (\mathbb{R}^{n_y} \times \mathcal{P})$ has a simple representation, namely $\mathcal{C} \cap (\mathbb{R}^{n_y} \times \mathcal{P}_r)$.

Theorem 2.2. *Consider a set of integer vectors $\mathcal{P} \subseteq \mathbb{Z}_+^{n_z}$. Let $\mathcal{P}_r \subseteq \mathbb{R}_+^{n_z}$ be as defined in* (2.5)*, and let $\mathcal{C} \subseteq \mathbb{R}^{n_y} \times \mathbb{R}^{n_z}$ be as defined in* (2.4)*, with n_a being the number of rows of A. Assume w.l.o.g. that $\begin{bmatrix} -A & B \\ A & -D \end{bmatrix}$ has linearly independent rows. Consider the following set:*

$$\mathcal{F}_2 = \mathcal{C} \cap (\mathbb{R}^{n_y} \times \mathcal{P}) = \{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathcal{P} \mid B\boldsymbol{z} \leq A\boldsymbol{y} \leq D\boldsymbol{z}\}.$$

If $\text{conv}(\mathcal{P}) = \mathcal{P}_r$, if $n_a \leq n_y$, and if each entry of $D - B$ is nonnegative, then the convex hull of $\mathcal{F}_2$ is:

$$\mathcal{G}_2 = \mathcal{C} \cap (\mathbb{R}^{n_y} \times \mathcal{P}_r) = \{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid B\boldsymbol{z} \leq A\boldsymbol{y} \leq D\boldsymbol{z}, H\boldsymbol{z} \leq \boldsymbol{h}, \boldsymbol{z} \geq \boldsymbol{0}\}.$$

Proof. Because

$$\mathcal{F}_2 = \{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{Z}^{n_z} \mid B\boldsymbol{z} \leq A\boldsymbol{y} \leq D\boldsymbol{z}, H\boldsymbol{z} \leq \boldsymbol{h}, \boldsymbol{z} \geq \boldsymbol{0}\},$$

it suffices to show that any such extreme point $(\boldsymbol{y}^*, \boldsymbol{z}^*)$ of $\mathcal{G}_2$ has an integral $\boldsymbol{z}^* \in \mathcal{P}$ [14, Theorem 4.3].

By definition, $(\boldsymbol{y}^*, \boldsymbol{z}^*)$ is the intersection of $n_y + n_z$ linearly independent hyperplanes. That is, among the inequalities that define $\mathcal{G}_2$, there exist $n_y + n_z$ of them that 1) have linearly independent coefficient vectors; 2) are active at $(\boldsymbol{y}^*, \boldsymbol{z}^*)$. Only $B\boldsymbol{z} \leq A\boldsymbol{y} \leq D\boldsymbol{z}$ involve $\boldsymbol{y}$; at least n_y of the $n_y + n_z$ linearly independent active inequalities must be from them. There are two possibilities: 1) none of the double-sided inequalities is active on both sides; and, 2) at least one double-sided inequality is active on both sides. We prove the integrality of $\boldsymbol{z}^*$ by considering these two cases separately.

Case 1). Since $n_a \leq n_y$, and since the double-sided inequalities can only be active on one side, at most n_y linearly independent hyperplanes can be from the double-sided inequalities. Therefore, out of the $n_y + n_z$ linearly independent hyperplanes, exactly n_y of them are from $B\boldsymbol{z} \leq A\boldsymbol{y} \leq D\boldsymbol{z}$, and n_z of them are from $H\boldsymbol{z} \leq \boldsymbol{h}$ and $\boldsymbol{z} \geq \boldsymbol{0}$. We collect the coefficients of the n_y active inequalities into matrix $A_= \in \mathbb{R}^{n_y \times n_y}$ and $G \in \mathbb{R}^{n_y \times n_z}$, so that $A_= \boldsymbol{y}^* = G\boldsymbol{z}^*$.[4]

[4]The rows of G may come from either B or D, depending on which side of the inequality is active.

We next show that the projection of

$$\{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid A_{=}\boldsymbol{y} = G\boldsymbol{z}\}$$

onto the $\boldsymbol{z}$-space is the whole of $\boldsymbol{z} \in \mathbb{R}^{n_z}$. To see this, suppose we are given any $\boldsymbol{z}' \in \mathbb{R}^{n_z}$; the point $(A_{=}^{-1}G\boldsymbol{z}', \boldsymbol{z}')$ lies in the set (2.2.2).[5]

Consequently, the value of $\boldsymbol{z}^*$ is completely determined by the n_z linearly independent active inequalities from $H\boldsymbol{z} \leq \boldsymbol{h}$ and $\boldsymbol{z} \geq \boldsymbol{0}$. That is to say, $\boldsymbol{z}^*$ is an extreme point of $\text{conv}(\mathcal{P})$. Since $\text{conv}(\mathcal{P}) = \mathcal{P}_r$, $\boldsymbol{z}^* \in \mathcal{P}$.

Case 2). Let the non-empty set $\mathcal{K} \subseteq \{1, \ldots, n_a\}$ be the set of indices for the double-sided inequalities that are active on both sides. For each $k \in \mathcal{K}$,

$$B_k\boldsymbol{z} = A_k\boldsymbol{y} = D_k\boldsymbol{z}, \tag{2.6}$$

where A_k, B_k and D_k are respectively the k-th row of A, B and D. In this case, the system of linear equations that define $(\boldsymbol{y}^*, \boldsymbol{z}^*)$ has degeneracy. Therefore, we first simplify (2.6).

The set

$$\{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid B_k\boldsymbol{z} = A_k\boldsymbol{y} = D_k\boldsymbol{z}\}$$

is equivalent to

$$\{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid A_k\boldsymbol{y} = D_k\boldsymbol{z}, B_k\boldsymbol{z} = D_k\boldsymbol{z}\},$$

[5]Inverting $A_{=}$ is justified by its linearly independent rows, which is implied by the linear independence of the rows of $\begin{bmatrix} -A & B \\ A & -D \end{bmatrix}$.

which is equivalent to

$$\{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid A_k\boldsymbol{y} = D_k\boldsymbol{z}, \\ z_i = 0, \forall i \text{ such that } D_{ki} - B_{ki} > 0\}.$$

To see this, recall the nonnegativity of $D_k - B_k$ and $\boldsymbol{z}$. The equation $(D_k - B_k)\boldsymbol{z} = \sum_i [(D_{ki} - B_{ki})z_i] = 0$ implies that for each i:

- if $D_{ki} - B_{ki} > 0$, then $z_i = 0$;
- if $D_{ki} - B_{ki} = 0$, then any $z_i \geq 0$ satisfies the equation.

Therefore, the set

$$\{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid B_k\boldsymbol{z} = A_k\boldsymbol{y} = D_k\boldsymbol{z}, \forall k \in \mathcal{K}\}$$

is equal to

$$\{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid A_k\boldsymbol{y} = D_k\boldsymbol{z}, \forall k \in \mathcal{K}, \\ z_i = 0, \forall i \text{ such that} \\ \exists k \in \mathcal{K} \text{ for which } D_{ki} - B_{ki} > 0\}. \tag{2.7}$$

After the simplification, only $|\mathcal{K}|$ linearly independent equalities involve $\boldsymbol{y}$ in (2.7). At the extreme point $(\boldsymbol{y}^*, \boldsymbol{z}^*)$, there still need to exist $n_y - |\mathcal{K}|$ additional active inequalities that involve $\boldsymbol{y}$; each of these active inequalities can only be from a double-sided inequality that is active on one side.

We collect the $\boldsymbol{y}$-coefficients and $\boldsymbol{z}$-coefficients of active inequalities in $\mathcal{K}$ into matrix $A_2 \in \mathbb{R}^{|\mathcal{K}| \times n_y}$ and $D_2 \in \mathbb{R}^{|\mathcal{K}| \times n_y}$, respectively. We also collect the $\boldsymbol{y}$-coefficients and the $\boldsymbol{z}$-coefficients of the other $n_y - |\mathcal{K}|$ active inequalities

involving $\boldsymbol{y}$ into matrices $A_1 \in \mathbb{R}^{(n_y-|\mathcal{K}|)\times n_y}$ and $E \in \mathbb{R}^{(n_y-|\mathcal{K}|)\times n_z}$,[6] respectively. As a result, at the extreme point $(\boldsymbol{y}^*, \boldsymbol{z}^*)$, we have the following active inequalities that involve $\boldsymbol{y}$:

$$\begin{bmatrix} A_2 \\ A_1 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} D_2 \\ E \end{bmatrix} \boldsymbol{z} \tag{2.8}$$

Similar to Case 1), the projection of

$$\{\boldsymbol{y} \in \mathbb{R}^{n_y}, \boldsymbol{z} \in \mathbb{R}^{n_z} \mid (2.8)\}$$

onto the $\boldsymbol{z}$-space is the whole of $\boldsymbol{z} \in \mathbb{R}^{n_z}$. Therefore, the value of $\boldsymbol{z}^*$ is determined by n_z linearly independent hyperplanes from the following set:

$$\begin{aligned} \{\boldsymbol{z} \in \mathbb{R}^{n_z} \mid & H\boldsymbol{z} \leq \boldsymbol{h}, \boldsymbol{z} \geq \boldsymbol{0}, \\ & z_i = 0, \forall i \text{ such that } \exists k \in \mathcal{K} \text{ for which } B_{ki} > 0\}. \end{aligned} \tag{2.9}$$

That is, $\boldsymbol{z}^*$ is an extreme point of the set (2.9).

Compared to Case 1), we have additional equalities from (2.7) that fix some entries of $\boldsymbol{z}$ to be zero. Because we already have the inequalities $\boldsymbol{z} \geq \boldsymbol{0}$ that define $\mathcal{P}_r$, the extreme points of the set (2.9) can only be a subset of the extreme points of $\mathcal{P}_r$. Together with the fact that $\text{conv}(\mathcal{P}) = \mathcal{P}_r$, we conclude that $\boldsymbol{z}^* \in \mathcal{P}$. □

We note that the proof of Theorem 2.2 still works if not all of inequalities $B\boldsymbol{z} \leq A\boldsymbol{y} \leq D\boldsymbol{z}$ are double-sided.

[6]Each row of E may be from either a row of B or a row of D, depending on which side of the inequality is active.

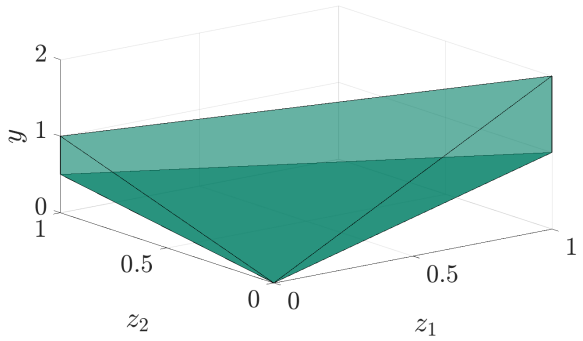


Figure 2.2: An example of a polyhedral cone for which Theorem 2.2 applies. The convex hull of this mixed-integer set is the intersection of the polyhedral cone $\mathcal{C}$ and the triangular prism $(\mathbb{R} \times \mathcal{P}_r)$.

Fig. 2.2 shows an example for which Theorem 2.2 applies. In this example, $n_y = 1$ and $n_z = 2$. The mixed-integer set is expressed as:

$$\{(y \in \mathbb{R}, \boldsymbol{z} \in \{0,1\}^2) \,|\, z_1 + 0.5z_2 \leq y \leq 2z_1 + z_2, \\ z_1 + z_2 \leq 1\},$$

for which

$$\mathcal{C} = \{(y \in \mathbb{R}, \boldsymbol{z} \in \mathbb{R}^2) \,|\, z_1 + 0.5z_2 \leq y \leq 2z_1 + z_2\}$$

is a polyhedral cone that is the intersection of two half-spaces, and

$$\mathcal{P}_r = \{\boldsymbol{z} \in \mathbb{R}^2 \,|\, z_1 + z_2 \leq 1, z_1 \geq 0, z_2 \geq 0\}$$

is an integral triangle. For this example, $n_a = n_y = 1$, and $D - B = [1 \quad 0.5]$ is nonnegative. Therefore, Theorem 2.2 implies that the convex hull of the above set is simply the intersection of $\mathbb{R} \times \mathcal{P}_r$ and $\mathcal{C}$.

Fig. 2.3 shows an example for which Theorem 2.2 does not apply. In this example, $n_1 = 1$ and $n_z = 2$. The mixed-integer set is:

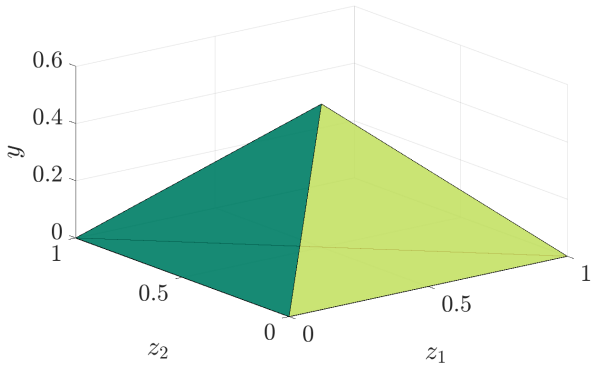


Figure 2.3: An example of a polyhedral cone for which Theorem 2.2 does not apply. The tetrahedron shown in this figure is the intersection of the polyhedral cone $\mathcal{C}$ and the triangular prism $\mathcal{P}_r$. It is not the convex hull of the mixed-integer set, because $\mathcal{C} \cap (\mathbb{R} \times \mathcal{P}_r)$ has a fractional extreme point $(0.5, 0.5, 0.5)$.

$$\begin{aligned}\{(y \in \mathbb{R}, \boldsymbol{z} \in \{0,1\}^2) \,|\, & 0 \leq y \leq z_1, \\ & y \leq z_2, \\ & z_1 + z_2 \leq 1\},\end{aligned}$$

for which

$$\mathcal{C} = \{(y \in \mathbb{R}, \boldsymbol{z} \in \mathbb{R}^2) \,|\, 0 \leq y \leq z_1, y \leq z_2\}$$

is a polyhedral cone, and

$$\mathcal{P}_r = \{\boldsymbol{z} \in \mathbb{R}^2 \,|\, z_1 + z_2 \leq 1, z_1 \geq 0, z_2 \geq 0\}$$

is the same triangle as in the previous example.

In this example, we have one double-sided inequality, and an additional inequality that involves y, which precludes the application of Theorem 2.2. The intersection $\mathcal{C} \cap (\mathbb{R} \times \mathcal{P}_r)$ contains a fractional vertex $(0.5, 0.5, 0.5)$, where the

two linearly independent inequalities involving y are active. The convex hull of this mixed-integer set is not $\mathcal{C} \cap (\mathbb{R} \times \mathcal{P}_r)$, but $\{0\} \times \mathcal{P}_r$, the two-dimensional triangle on the $y = 0$ plane.

In the context of unit commitment, Theorem 2.2 applies to the feasible set of an individual generating unit as defined in the next section. We also utilize Theorem 2.2 in Chapter 5 when we study the feasible region of a combined cycle unit.

2.3 The Unit Commitment Problem and Its Lagrangian Dual

We consider a T-period offer-based UC problem with $|\mathcal{G}|$ units. For unit $g \in \mathcal{G}$ at time $t \in \{1, \ldots, T\}$, the commitment variable x_{gt} is 1 if the unit is online and is 0 if the unit is offline. The start-up variable u_{gt} is 1 if unit g starts up at period t and is 0 otherwise.

Denote unit g's dispatch-level (dispatched power output) vector by $\boldsymbol{p}_g \in \mathbb{R}^T_+$, whose t-th component is the dispatch level at time t. Similarly, $\boldsymbol{x}_g \in \{0,1\}^T$ denotes the commitment vector, and $\boldsymbol{u}_g \in \{0,1\}^{T-1}$ is the start-up vector.[7] Let unit g's offer cost function be $C_g(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g)$, which may include energy, start-up, and no-load costs. As in [30, 59], we assume that C_g is convex piecewise linear or convex quadratic in $\boldsymbol{p}_g$. We assume that the start-

[7] The start-up variables $\boldsymbol{u}_g$ are defined from period 2 to T. For simplicity we do not consider the initial conditions of the units. Our formulation can be extended to consider these conditions.

up and no-load costs are independent of time t and independent of the time since the last shut-down. Let $\mathcal{X}_g \subseteq \mathbb{R}_+^T \times \{0,1\}^T \times \{0,1\}^{T-1}$ be the set of feasible commitment and dispatch decisions for unit g. We assume that private constraints that define $\mathcal{X}_g$ are specified by linear inequalities: these constraints may include generation limits, minimum up/down time, and perhaps ramping constraints [30]. For simplicity, we first consider a UC problem in which the only type of system-wide constraint is the power balance constraint.

Let $\boldsymbol{d} \in \mathbb{R}_+^T$ be a demand vector, whose t-th component denotes system demand at time t. Since we do not consider elastic demand for simplicity, maximizing social welfare is equivalent to minimizing total operating costs. The UC problem makes a set of commitment and dispatch decisions that minimizes the total cost, while satisfying physical and operational constraints:

$$
\begin{aligned}
v(\boldsymbol{d}) = \quad & \min_{\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g,\, g \in \mathcal{G}} && \sum_{g \in \mathcal{G}} C_g(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) && (2.10)\\
& \text{s.t.} && \sum_{g \in \mathcal{G}} \boldsymbol{p}_g = \boldsymbol{d} && (2.11)\\
& && (\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) \in \mathcal{X}_g \quad \forall g \in \mathcal{G}. && (2.12)
\end{aligned}
$$

In addition, we view the UC problem as parametrized by the demand vector $\boldsymbol{d}$, and denote the value function of the UC problem by $v(\boldsymbol{d})$.

In Lagrangian relaxation [43], we dualize all system-wide constraints. Since the only type of system-wide constraints we consider for now is power balance constraints (2.11), we dualize these constraints and obtain the La-

grangian dual function:

$$q(\boldsymbol{\pi}) = \sum_{g \in \mathcal{G}} \left(\min_{(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) \in \mathcal{X}_g} C_g(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) - \boldsymbol{\pi}^\intercal \boldsymbol{p}_g \right) + \boldsymbol{\pi}^\intercal \boldsymbol{d}, \tag{2.13}$$

where $\boldsymbol{\pi} \in \mathbb{R}^T$ is now the dual vector associated with the power balance constraints. The Lagrangian dual problem is:

$$\max_{\boldsymbol{\pi}} \quad q(\boldsymbol{\pi}). \tag{2.14}$$

The Langrangian dual problem is convex but non-smooth. Algorithms such as sub-gradient methods, bundle methods [43], and cutting plane methods [27] have been proposed to solve the Lagrangian dual problem of mixed-integer programming problems. None of the above-mentioned methods guarantees convergence in polynomial time. For non-smooth optimization techniques like the sub-gradient method for which no certificate of optimality exists, the algorithm is often terminated before an optimal value is attained [86, Section 10.3]. Efficient computation of the Lagrangian dual problem remains challenging.

2.4 A Primal Formulation of the Lagrangian Dual

This section proposes a primal formulation for the Lagrangian dual problem of the UC problem. For each unit in this formulation, the feasible set is replaced by its convex hull, and the cost function is replaced by its convex envelope.

2.4.1 A Primal Formulation for the Lagrangian Dual Problem

We make the following:

Assumption 2.1. The set $\mathcal{X}_g$ is compact for all $g \in \mathcal{G}$, and all system-wide constraints are linear.

Let $C^{**}_{g,\mathcal{X}_g}(\cdot)$ be the convex envelope of $C_g(\cdot)$ *taken over* $\mathcal{X}_g$. The function $C^{**}_{g,\mathcal{X}_g}(\cdot)$ is the largest convex function on $\text{conv}(\mathcal{X}_g)$ that is an under-estimator of C_g on $\mathcal{X}_g$. It is also the conjugate of the conjugate of C_g.

Note that the UC problem is separable across g absent the system-wide constraints. We have [37, Theorem 1]:

Theorem 2.3. *Under Assumption 2.1, (a) the optimal objective function value of the Lagrangian dual problem* (2.14) *equals the minimum of the following problem denoted by* **LD-Primal***:*

$$\min_{\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g, g \in \mathcal{G}} \quad \sum_{g \in \mathcal{G}} C^{**}_{g,\mathcal{X}_g}(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) \tag{2.15}$$

$$\text{s.t.} \quad \sum_{g \in \mathcal{G}} \boldsymbol{p}_g = \boldsymbol{d} \tag{2.16}$$

$$(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) \in \text{conv}(\mathcal{X}_g) \quad \forall g \in \mathcal{G}, \tag{2.17}$$

and (b) an optimal dual vector associated with (2.16) *is an optimal solution to* (2.14).

Proof. Since the private and system-wide constraints are defined by linear equalities and inequalities, strong duality holds between **LD-Primal** and its

Lagrangian dual problem. Therefore, Theorem 2.3 holds by Theorem 3.3 in [23]. □

Theorem 2.3 suggests that if we have an explicit characterization of $C^{**}_{g,\mathcal{X}_g}(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g)$ and $\text{conv}(\mathcal{X}_g)$, we can solve **LD-Primal** to obtain the dual maximizers of (2.14).

2.4.2 Characterization of the Convex Hulls

As mentioned in Section 2.2, describing the convex hull of an arbitrary non-convex set is difficult in general. A general-purpose method proposed in [4] is used in [75] to obtain a convex hull description of a unit's feasible set. This method applies to a feasible set defined by arbitrary linear constraints. All feasible commitment decisions are enumerated in this method, and both the number of variables and number of constraints in the resulting description are exponential in the number of time periods.

A recent polyhedral study of a unit's state-transition polytope [69] exploits its structure. Using these special-purpose valid inequalities that characterize the convex hull of the feasible binary commitment variables, we apply the theorems in Section 2.2 to obtain a tractable description of $\text{conv}(\mathcal{X}_g)$, the convex hull of the feasible region of an individual unit.

Once committed, many generators must remain on for a minimum up time. Similarly, once decommitted, many generators must remain offline for a minimum down time. Let L_g and l_g, respectively, be the minimum up and

minimum down times for unit g, and let $\underline{p}_g$ and $\overline{p}_g$ be the minimum and maximum generation levels for unit g. We consider feasible commitment and dispatch decisions of a unit limited by:

- state-transition constraints that represent the relationship between binary variables:

$$u_{gt} \geq x_{gt} - x_{g,t-1}, \quad \forall t \in [2, T], \tag{2.18}$$

- minimum up/down time constraints using the formulation in [69]:

$$\sum_{i=t-L_g+1}^{t} u_{gi} \leq x_{gt}, \qquad \forall t \in [L_g + 1, T], \tag{2.19}$$

$$\sum_{i=t-l_g+1}^{t} u_{gi} \leq 1 - x_{g,t-l_g}, \quad \forall t \in [l_g + 1, T], \tag{2.20}$$

- dispatch level limits:

$$x_{gt}\underline{p}_g \leq p_{gt} \leq x_{gt}\overline{p}_g, \quad \forall t \in [1, T]. \tag{2.21}$$

Ramping constraints are not considered until Section 2.5.3. Therefore, a unit's feasible set is

$$\mathcal{X}_g = \{\boldsymbol{p}_g \in \mathbb{R}^T, \boldsymbol{x}_g \in \{0,1\}^T, \boldsymbol{u}_g \in \{0,1\}^{T-1} \,|\, (2.18)\text{–}(2.20), (2.21)\} \tag{2.22}$$

The set of feasible binary decisions alone is:

$$\mathcal{D}_g = \{\boldsymbol{x}_g \in \{0,1\}^T, \boldsymbol{u}_g \in \{0,1\}^{T-1} \,|\, (2.18)\text{–}(2.20)\}. \tag{2.23}$$

The following trivial inequalities are valid for $\mathcal{D}_g$:

$$u_{gt} \geq 0, \quad \forall t \in [2, T]. \tag{2.24}$$

It is claimed in [69] that

$$\text{conv}(\mathcal{D}_g) = \{\boldsymbol{x}_g \in \mathbb{R}^T, \boldsymbol{u}_g \in \mathbb{R}^{T-1} \,|(2.18)\text{–}(2.20), (2.24)\}. \tag{2.25}$$

While this statement is true, the proof of it relies on a lemma that states that all the extreme points of the polytope defined in (2.25) are integral, Lemma 2.9 in [69]. The proof of Lemma 2.9 given in [69] is flawed, however. We identify the flaws and give our own proof in Appendix of this dissertation.

We also note that the minimum up/down time constraints in [69] have been used in recent literature on tight formulations of the UC problem, such as [52, 54, 59].

The following theorem extends the result in (2.25) to also include the dispatch decisions [37, Theorem 2]:

Theorem 2.4. *The result shown in* (2.25) *implies that*

$$\mathcal{C}_g = \{\boldsymbol{p}_g \in \mathbb{R}^T, \boldsymbol{x}_g \in \mathbb{R}^T, \boldsymbol{u}_g \in \mathbb{R}^{T-1} \,|\, (2.18)\text{–}(2.21), (2.24)\}$$

describes the convex hull of $\mathcal{X}_g$.[8]

Proof. We show that this theorem is a special case of Theorem 2.2. In this special case, we have $\boldsymbol{z} = (\boldsymbol{x}_g, \boldsymbol{u}_g)$ and $\boldsymbol{y} = \boldsymbol{p}_g$. The set of feasible integer vectors $\mathcal{P}$ is described by $\mathcal{D}_g$, and we have its convex hull description $\mathcal{P}_r$ shown in (2.25). The constraints that couple $\boldsymbol{z}$ and $\boldsymbol{y}$ and define the polyhedral cone $\mathcal{C}$ are the dispatch limit constraints (2.21). In this case, we have $n_y = n_a = T$,

[8]For $T = 2$ and $T = 3$, the convex hull of a more general $\mathcal{X}_g$ with ramping constraints has been characterized in [64]. The result here does not consider ramping constraints, but holds for an arbitrary T.

$A = I$, $B = \text{diag}(\underline{p}_g)$, and $D = \text{diag}(\overline{p}_g)$. Since each entry of the matrix $D - B$ is nonnegative, applying Theorem 2.2 to this special case leads to desired results. □

2.4.3 Characterization of the Convex Envelopes

In **LD-Primal**, each cost function $C_g(\cdot)$ is replaced by its convex envelope taken over the non-convex feasible set $\mathcal{X}_g$. When a unit has a constant marginal cost, $C_g(\cdot)$ is affine, and the convex envelope of $C_g(\cdot)$ has the same functional form as $C_g(\cdot)$ itself.

When $C_g(\cdot)$ is not affine (piecewise linear or quadratic in $\boldsymbol{p}_g$), its convex envelope has a different functional form. We first discuss the convex quadratic case.

Let the start-up and no-load cost of unit g be h_g and c_g, respectively. Define the following set:

$$\mathcal{X}_{gt} = \{p_{gt} \in \mathbb{R}, x_{gt} \in \{0,1\}, u_{gt} \in \{0,1\} \mid x_{gt}\underline{p}_g \leq p_{gt} \leq x_{gt}\overline{p}_g\}.$$

Suppose the offer cost function $C_g : \mathcal{X}_g \to \mathbb{R}$ is defined by:

$$C_g(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) = \sum_{t=1}^{T} C_{gt}(p_{gt}, x_{gt}, u_{gt}), \tag{2.26}$$

where for each period t, $C_{gt} : \mathcal{X}_{gt} \to \mathbb{R}$ is a convex quadratic function defined by a *single-period cost function*:

$$C_{gt}(p_{gt}, x_{gt}, u_{gt}) = a_g p_{gt}^2 + b_g p_{gt} + c_g x_{gt} + h_g u_{gt}, \tag{2.27}$$

where we assume $a_g > 0$.

Theorem 2.5. *The convex envelope of the quadratic cost function C_g taken over $\mathcal{X}_g$ is the function $C^{**}_{g,\mathcal{X}_g} : \mathrm{conv}(\mathcal{X}_g) \to \mathbb{R}$ defined by the following:*

$$C^{**}_{g,\mathcal{X}_g}(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) = \sum_{t=1}^{T} C^{**}_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, u_{gt}),$$

where $C^{**}_{gt,\mathcal{X}_{gt}} : \mathrm{conv}(\mathcal{X}_{gt}) \to \mathbb{R}$ is defined by the following:

$$C^{**}_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, u_{gt}) = \begin{cases} a_g \frac{p_{gt}^2}{x_{gt}} + b_g p_{gt} + c_g x_{gt} + h_g u_{gt}, & x_{gt} > 0, \\ 0, & x_{gt} = 0. \end{cases}$$

Proof. We first show that $C^{**}_{gt,\mathcal{X}_{gt}}$ is the convex envelope of the single-period cost function C_{gt}. Consider the value of the function $C^{**}_{gt,\mathcal{X}_{gt}}$ with u_{gt} fixed at zero:

$$C^{**}_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, 0) = \begin{cases} a_g \frac{p_{gt}^2}{x_{gt}} + b_g p_{gt} + c_g x_{gt}, & x_{gt} > 0, \\ 0, & x_{gt} = 0, \end{cases} \tag{2.28}$$

as (p_{gt}, x_{gt}) varies over $\{p_{gt} \in \mathbb{R}, x_{gt} \in [0,1] \,|\, x_{gt}\underline{p}_g \le p_{gt} \le x_{gt}\overline{p}_g\}$. At points where $x_{gt} \in \{0,1\}$, the value of this function is the same as $C_{gt}(p_{gt}, x_{gt}, 0)$. The value of $C^{**}_{gt,\mathcal{X}_{gt}}$ at any point (p_{gt}, x_{gt}) with $x_{gt} \in (0,1)$ is determined by linear interpolation of $C_{gt}(p_{gt}, x_{gt}, 0)$ between $(0,0)$ and $(\frac{p_{gt}}{x_{gt}}, 1)$. The function $C^{**}_{gt,\mathcal{X}_{gt}}$ is continuous and convex on $\{p_{gt} \in \mathbb{R}, x_{gt} \in [0,1] \,|\, x_{gt}\underline{p}_g \le p_{gt} \le x_{gt}\overline{p}_g\}$, as can be verified by taking its Hessian in this domain.

We prove by contradiction that among the convex under-estimators of $C_{gt}(p_{gt}, x_{gt}, 0)$ on the given domain, $C^{**}_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, 0)$ is the largest one. Suppose not, then there exists a convex under-estimator of $C_{gt}(p_{gt}, x_{gt}, 0)$, denoted by $C'_{gt}(p_{gt}, x_{gt}, 0)$, for which there exist a point (p'_{gt}, x'_{gt}) with $x'_{gt} \in (0,1)$ such that $C'_{gt}(p'_{gt}, x'_{gt}, 0) > C^{**}_{gt}(p'_{gt}, x'_{gt}, 0)$. Consider the line interval connecting

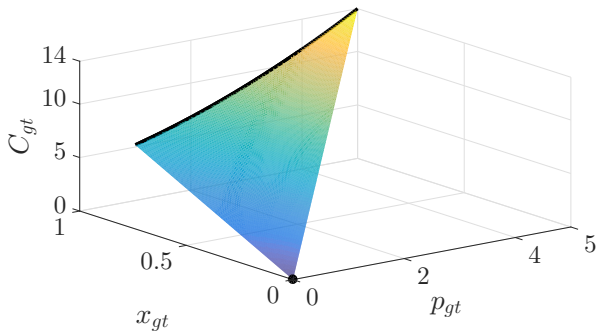


Figure 2.4: Convex envelope of a single-period cost function. The figure shows the graph of a single-period cost function $0.2p_{gt}^2 + p_{gt} + 4x_{gt}$ defined on $\{x_{gt} \in \{0,1\}, p_{gt} \in \mathbb{R} \,|\, x_{gt} \leq p_{gt} \leq 5x_{gt}\}$ (black dot and black curved line), together with its convex envelope (the colored surface).

$(0,0,0)$ and $(\frac{p'_{gt}}{x'_{gt}}, 1, C_{gt}(\frac{p'_{gt}}{x'_{gt}}, 1, 0))$. We have $C'_{gt}(p'_{gt}, x'_{gt}, 0) > C^{**}_{gt}(p'_{gt}, x'_{gt}, 0) = 0 + x'_{gt} C_{gt}(\frac{p'_{gt}}{x'_{gt}}, 1, 0)$, which implies that C'_{gt} is not convex when restricted to this line. This contradicts the convexity of C'_{gt}, since a function is convex if and only if it is convex when restricted to any line that intersects its domain [9, Chapter 3]. Therefore, $C^{**}_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, 0)$ is the convex envelope of $C_{gt}(p_{gt}, x_{gt}, 0)$.

Fig. 2.4 shows an example of $C_{gt}(p_{gt}, x_{gt}, 0)$ and its convex envelope. Since C_{gt} is affine in u_{gt}, $C^{**}_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, u_{gt})$ is the convex envelope of $C_{gt}(p_{gt}, x_{gt}, u_{gt})$.

Now consider the time-coupled domain $\mathcal{X}_g$. Let $\mathcal{X}'_g = \prod_{t=1}^{T} \mathcal{X}_{gt}$, so that $\mathcal{X}_g \subseteq \mathcal{X}'_g$. The set $\mathcal{X}'_g \backslash \mathcal{X}_g$ contains the commitment and dispatch decisions that satisfy generation limit constraints (2.21) but not constraints (2.18)–(2.20). Because constraints (2.18)–(2.20) restrict the binary variables $\boldsymbol{x}_g$ and $\boldsymbol{u}_g$, $(\mathcal{X}'_g \backslash \mathcal{X}_g) \cap \text{conv}(\mathcal{X}_g) = \emptyset$. Consequently, $\mathcal{X}'_g \cap \text{conv}(\mathcal{X}_g) = \mathcal{X}_g$.

By the separability of C_g and $\mathcal{X}'_g$ across t, we have that the convex envelope of C_g taken over $\mathcal{X}'_g$ is the function $C^{**}_{g,\mathcal{X}'_g} : \text{conv}(\mathcal{X}'_{\text{g}}) \to \mathbb{R}$ defined by:

$$C^{**}_{g,\mathcal{X}'_g}(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) = \sum_{t=1}^{T} \begin{cases} a_g \frac{p_{gt}^2}{x_{gt}} + b_g p_{gt} + c_g x_{gt} + h_g u_{gt}, & x_{gt} > 0, \\ 0, & x_{gt} = 0. \end{cases}$$

Note that taking the convex envelope of a function is equivalent to taking the convex hull of its epigraph. Because $\mathcal{X}'_g \cap \text{conv}(\mathcal{X}_g) = \mathcal{X}_g$, on $\text{conv}(\mathcal{X}_g)$, the convex hull of the epigraph of C_g taken over $\mathcal{X}_g$ is the same as the convex hull of the epigraph taken over $\mathcal{X}'_g$. Therefore, on $\text{conv}(\mathcal{X}_g)$, $C^{**}_{g,\mathcal{X}'_g}(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) = C^{**}_{g,\mathcal{X}_g}(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g)$. □

Using the convex envelopes gives us a better lower bound than simply keeping the functional form of $C_{gt,\mathcal{X}_{gt}}$ and relaxing its domain. Given $g \in \mathcal{G}$, $t \in [1, T]$, and $p_{gt} > 0$, when x_{gt} is integral, $C^{**}_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, 0) = C_{gt,\mathcal{X}_{gt}}(p_{gt}, x_{gt}, 0)$; however, when x_{gt} is fractional, we have $\frac{p_{gt}^2}{x_{gt}} > p_{gt}^2$, providing a tighter bound.

We next consider convex piecewise linear cost functions. Suppose the interval $[\underline{p}_g, \overline{p}_g]$ is partitioned into $|\mathcal{K}|$ intervals:

$$[\underline{p}_g, \overline{p}_g] = \bigcup_{k \in \mathcal{K}} \mathcal{I}_k, \tag{2.29}$$

where k is the index for the partitioned intervals.

Recalling that h_g denotes the start-up cost of unit g, suppose at each period t, when $p_{gt} \in \mathcal{I}_k$, the operating cost (excluding start-up cost and no-load cost) is

$$\tilde{C}_{gt}(p_{gt}, 1, 0) = a_{gk} p_{gt} + b_{gk}. \tag{2.30}$$

Introducing an auxiliary variable $q_{gt} \in \mathbb{R}_+$, the single-period piecewise linear cost function can be implemented as

$$\tilde{C}_{gt}(p_{gt}, x_{gt}, u_{gt}) = q_{gt} + c_g x_{gt} + h_g u_{gt}, \tag{2.31}$$

$$q_{gt} \geq a_{gk} p_{gt} + b_{gk}, \forall k \in \mathcal{K}. \tag{2.32}$$

The convex envelope of a convex piecewise linear cost function taken over $\mathcal{X}_g$ is also piecewise linear. The proof is similar to that for Theorem 2.5.

Theorem 2.6. *The convex envelope of the convex piecewise linear cost function* $\tilde{C}_{gt}$ *defined in* (2.31) *taken over* $\mathcal{X}_g$ *is the function* $\tilde{C}^{**}_{g,\mathcal{X}_g} : \text{conv}(\mathcal{X}_g) \to \mathbb{R}$ *defined by:*

$$\tilde{C}^{**}_{g,\mathcal{X}_g}(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) = \sum_{t=1}^{T} \begin{cases} a_{gk} p_{gt} + (c_g + b_{gk}) x_{gt} + h_g u_{gt}, & \text{if } \frac{p_{gt}}{x_{gt}} \in \mathcal{I}_k, \\ 0, & \text{if } x_{gt} = 0. \end{cases}$$

Typically, b_{gk} is non-positive. Therefore, similar to the quadratic case, using the convex envelopes gives us a better lower bound than the implementation shown in (2.31) and (2.32).

2.4.4 Reformulation and Polynomial-Time Solution

When the cost functions are quadratic, non-linearity of the convex envelope comes from the quadratic-over-linear terms $a_g \frac{p_{gt}^2}{x_{gt}}$, which are known to be convex. Moreover, we can move these terms from the objective into constraints and cast **LD-Primal** as a second-order cone program (SOCP).

For each g and t, we replace $a_g \frac{p_{gt}^2}{x_{gt}}$ by a new variable $s_{gt} \in \mathbb{R}_+$ and introduce the following constraint:

$$s_{gt} x_{gt} \geq a_g p_{gt}^2. \tag{2.33}$$

Since we are minimizing, when $x_{gt} = 0$, the optimal value for s_{gt} is zero, which is consistent with the convex envelope. For $x_{gt} \geq 0$ and $s_{gt} \geq 0$, constraint (2.33) is equivalent to

$$\|(2\sqrt{a_g} p_{gt}, x_{gt} - s_{gt})\|_2 \leq x_{gt} + s_{gt}, \tag{2.34}$$

which is a second-order cone constraint [47]. With this reformulation technique, **LD-Primal** can be cast as an SOCP, which can be solved in polynomial time using off-the-shelf interior-point solvers, e.g. GUROBI [31].

In the case where the cost functions are piecewise linear, the convex envelope of the cost function is convex piecewise linear. The resulting **LD-Primal** is a linear program (LP).

Since the number of constraints in our explicit formulation is polynomial in T and $|\mathcal{G}|$, **LD-Primal** can be solved as a convex program in polynomial time in both cases.

2.5 Extensions

2.5.1 Transmission and Other Linear System-Wide Constraints

In Lagrangian relaxation, all system-wide constraints are dualized. Theorem 2.3 applies to any linear system-wide constraints, but does not apply to nonlinear system-wide constraints [42].

We consider a linear approximation to the transmission constraints and augment **LD-Primal** with angle-eliminated transmission constraints in terms of the shift factors. Other linear system-wide constraints (e.g., contingency constraints, constraints that approximate loss in the transmission system) can be treated in a similar fashion.

2.5.2 Ancillary Services

When we co-optimize energy and ancillary services, we introduce a set of variables to represent the ancillary services provided by market participants. We use spinning reserve as an example.

Let vector $\boldsymbol{r}_g$ denote the amount of spinning reserve provided by generator g in each time period. Let $\underline{r}_g$ and $\overline{r}_g$ be the lower and upper limits on spinning reserve. We include the following constraints:

$$x_{gt}\underline{p}_g \leq p_{gt} + r_{gt} \leq x_{gt}\overline{p}_g, \quad \forall t \in [1, T], \tag{2.35}$$

$$x_{gt}\underline{r}_g \leq r_{gt} \leq x_{gt}\overline{r}_g, \quad \forall t \in [1, T]. \tag{2.36}$$

The feasible set for each unit is redefined to be:

$$\mathcal{X}_g = \{\boldsymbol{p}_g \in \mathbb{R}^T, \boldsymbol{r}_g \in \mathbb{R}^T, \boldsymbol{x}_g \in \{0,1\}^T, \boldsymbol{u}_g \in \{0,1\}^{T-1} \mid (2.18)\text{–}(2.20), (2.35), (2.36)\}.$$

We can show that:

$$\text{conv}(\mathcal{X}_g) = \{\boldsymbol{p}_g \in \mathbb{R}^T, \boldsymbol{r}_g \in \mathbb{R}^T, \boldsymbol{x}_g \in \mathbb{R}^T, \boldsymbol{u}_g \in \mathbb{R}^{T-1} \mid (2.18)\text{–}(2.20), (2.24), (2.35), (2.36)\}.$$

The proof is similar to that for Theorem 2.4.

Since the convex envelope of a convex function taken over a convex domain is the convex function itself, and since the additional constraints (2.35) and (2.36) define a convex feasible set, introducing ancillary services does not alter the convex envelope.

2.5.3 Ramping Constraints

Ramping constraints are a set of private constraints that limit the increase or decrease of power output from one time period to the next. Let $\overline{v}_g$ denote unit g's start-up/shut-down ramp rate limit, and let v_g be unit g's ramp-up/down rate when committed. The ramping constraints are [64]:

$$p_{gt} - p_{g,t-1} \leq v_g x_{g,t-1} + \overline{v}_g(1 - x_{g,t-1}),\ \forall t \in [2, T], \tag{2.37}$$

$$p_{g,t-1} - p_{gt} \leq v_g x_{gt} + \overline{v}_g(1 - x_{gt}),\ \forall t \in [2, T]. \tag{2.38}$$

We redefine the feasible set for each unit to be

$$\mathcal{X}_g = \{\boldsymbol{p}_g \in \mathbb{R}^T, \boldsymbol{x}_g \in \{0,1\}^T, \boldsymbol{u}_g \in \{0,1\}^{T-1} \,|\, (2.18)\text{–}(2.21), (2.37), (2.38)\}.$$

Ramping constraints define a convex feasible set, and thus do not change the convex envelope of the cost function. They complicate the convex hulls, however. When these time-coupled constraints are included in the definition of $\mathcal{X}_g$, equations (2.18)–(2.21), (2.24), together with ramping constraints

(2.37) and (2.38) themselves, do not completely characterize conv($\mathcal{X}_g$). Additional valid inequalities are needed to describe the convex hulls. The number of valid inequalities needed is in general exponential in T [18].

Explicit descriptions of conv($\mathcal{X}_g$) for the case where $T = 2$ are obtained in [64]. The following valid inequalities

$$p_{g,t-1} \leq \overline{v}_g x_{g,t-1} + (\overline{p}_g - \overline{v}_g)(x_{gt} - u_{gt}),\ \forall t \in [2, T] \tag{2.39}$$
$$p_{gt} \leq \overline{p}_g x_{gt} - (\overline{p}_g - \overline{v}_g) u_{gt},\ \forall t \in [2, T] \tag{2.40}$$
$$p_{gt} - p_{g,t-1} \leq (\underline{p}_g + v_g) x_{gt} - \underline{p}_g x_{g,t-1} - (\underline{p}_g + v_g - \overline{v}_g) u_{gt},\ \forall t \in [2, T] \tag{2.41}$$
$$p_{g,t-1} - p_{gt} \leq \overline{v}_g x_{g,t-1} - (\overline{v}_g - v_g) x_{gt} - (\underline{p}_g + v_g - \overline{v}_g) u_{gt},\ \forall t \in [2, T] \tag{2.42}$$

along with equations (2.18)–(2.21), (2.24), describe conv($\mathcal{X}_g$) for the case of $T = 2$. Explicit descriptions of conv($\mathcal{X}_g$) for $T = 3$ are also shown in [64]. More importantly, valid inequalities in these descriptions can be applied to any two or three consecutive time periods to tighten the approximation of conv($\mathcal{X}_g$) for $T > 3$.

When considering ramping constraints, we can solve an approximation of **LD-Primal** that includes the above-mentioned valid constraints for $T = 2$ and $T = 3$. Our description of conv($\mathcal{X}_g$) is not exact. Consequently, our approximation provides a lower bound for **LD-Primal**. The gap between the approximated **LD-Primal** and the UC problem gives an upper bound for the duality gap between the UC problem and its Lagrangian dual problem. In the next chapter, we apply **LD-Primal** to the problem of convex hull pricing. Empirical characteristics of this gap will also be explored.

Chapter 3

Convex Hull Pricing

3.1 Introduction

Some electricity markets are non-convex due to indivisibility of generators. Start-up and no-load costs of units may not be covered by sales of energy at locational marginal prices (LMPs). Consequently, the ISO has to make "uplift" payments to units that have an incentive to deviate from the ISO's solution. These uplift payments are detrimental to market transparency.

Transparency of the market can be improved by keeping uplift payments as low as possible. To this end, several alternative pricing schemes have been proposed. For example, an ad hoc method to reduce uplift payments to fast-start units is to relax their minimum generation limits to zero, so that they can set the LMPs [26]. Keeping marginal prices as uniform energy prices, the pricing scheme proposed in [57] introduces artificial constraints that set the commitment variables at a welfare-maximizing solution and create discriminatory side payments for commitment decisions based on the optimal dual variables associated with these artificial constraints. Different from the pricing

This chapter is based in part on the following publication to which the coauthors contributed equally: Bowen Hua and Ross Baldick. "A convex primal formulation for convex hull pricing." *IEEE Transactions on Power Systems* 32.5 (2017): 3814-3823.

schemes that aim at supporting a welfare-maximizing solution, pricing schemes such as [72] have been proposed to incentivize a commitment and dispatch solution that is close, but not necessarily equal to the ISO's welfare-maximizing solution. In these methods, allocative efficiency is traded off against transparency. Instead of focusing on unit-commitment based markets that are typical in the US, Ortner and Huppmann [58] define a quasi-equilibrium for a self-committed electricity market, and determine prices through a mathematical program with equilibrium constraints. See [45] for a comprehensive review of different pricing schemes for markets with non-convexities.

Convex hull pricing [29, 34, 70] is a pricing scheme that minimizes certain uplift payments over all possible uniform prices, and has received much attention. The Midcontinent ISO (MISO) has implemented an approximation of convex hull pricing, and refers to convex hull prices as extended locational marginal prices (ELMPs) [82]. Convex hull prices are slopes of the convex envelope of the system cost as a function of demand,[1] and are thus non-decreasing with respect to demand. These prices minimize the total uplift payment defined by the duality gap between the UC problem and its Lagrangian dual.

Convex hull prices are the dual maximizers of the Lagrangian dual of UC. As discussed in Section 2.1, Lagrangian dual problem is convex but non-smooth. In the context of convex hull pricing, the focus is on the opti-

[1] More precisely, they are sub-gradients of the convex envelope of the system cost function. The system cost function here is the value function of the UC problem parametrized by demand. The convex envelope of a function is the largest convex under-estimator of the given function.

mality of dual variables, rather than obtaining primal solutions. Therefore, in addition to general-purpose methods, an outer approximation method [29], a sub-gradient simplex cutting plane method [83], and an extreme-point subdifferential method [84] have been designed specifically for convex hull pricing.

None of the general-purpose non-smooth optimization methods or the above-mentioned special-purposed methods guarantees convergence in polynomial time. Obtaining exact convex hull prices has previously been thought to be computationally expensive [33, 82]. Consequently, MISO implements a single-period approximation of convex hull pricing that is based on a version of the UC problem in which integer variables are relaxed to being continuous [82].

Section 3.2 introduces the concept of uplift payment in non-convex electricity markets and gives the mathematical definition of convex hull prices. Section 3.3 shows that, utilizing the primal formulation for Lagrangian dual in Section 2, we can solve for exact convex hull pricing in polynomial time when ramping constraints are not present. A tight and tractable approximation is available when there are ramping constraint. Numerical results follow in Section 3.4.

3.2 Uplift Payments and Convex Hull Prices

Suppose that an energy price vector $\boldsymbol{\pi}$ is specified by the ISO. Assume that unit g is a price-taker.[2] Given price $\boldsymbol{\pi}$, its profit maximization problem

[2]We assume that the units are prices takers in the economic sense that they cannot affect prices.

is:

$$w_g(\boldsymbol{\pi}) = \max_{\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g} \quad \boldsymbol{\pi}^\intercal \boldsymbol{p}_g - C_g(\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) \tag{3.1}$$

$$\text{s.t.} \quad (\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g) \in \mathcal{X}_g, \tag{3.2}$$

where $w_g(\boldsymbol{\pi})$ is the value function of this problem.

From a microeconomic viewpoint, the ISO's UC problem (2.10)–(2.12) is a social planner's problem whose solution is welfare-maximizing. Problem (3.1)–(3.2) is the profit maximization problem of a rational agent. If the ISO's problem is convex and satisfies strong duality, we can set $\boldsymbol{\pi}$ to be the optimal dual vector associated with the supply-demand balance constraints in the planner's problem. As a result, there exist individual profit-maximizing decisions that align with the welfare-maximizing solution.[3] If the ISO's problem were strictly convex, then prices alone would provide sufficient information for the units to determine an efficient decision [49, Chapter 16].

However, the UC problem is non-convex because of the integer decision variables $\boldsymbol{x}_g$ and $\boldsymbol{u}_g$. Thus, in general, there does not exist a set of prices that support the ISO's decisions. In particular, revenues from locational marginal prices (LMPs) may not cover the offered costs of a unit, and the unit may prefer to deviate from the ISO's commitment and dispatch decisions, unless there is additional incentive to follow those decisions. LMPs are determined as the optimal dual variables associated with the supply-demand balance constraints

[3]The ISO may need to specify which solution is welfare maximizing in the presence of multiple profit-maximizing solutions to the individual problems.

in a continuous convex economic dispatch problem with commitment decisions fixed at ISO-determined optimal values.

One way to address the above-mentioned problem is for the ISO to maintain uniform energy prices and provide side payments to units whose individually rational decision is different from the ISO's. In principle, these side payments, also known as "uplift," should cover the gap between the maximum possible profit (the optimal objective function value of problem (3.1)–(3.2)), and the profit made by following the ISO's decision (the profit (3.1) evaluated at the ISO's decisions).

Mathematically, given a set of uniform energy prices $\boldsymbol{\pi}$ and an ISO's welfare-maximizing decision $(\boldsymbol{p}^*, \boldsymbol{x}^*, \boldsymbol{u}^*)$, the amount of uplift payment needed for unit g equals its *lost opportunity cost*:

$$U_g(\boldsymbol{\pi}, \boldsymbol{p}^*, \boldsymbol{x}^*, \boldsymbol{u}^*) = w_g(\boldsymbol{\pi}) - (\boldsymbol{\pi}^\intercal \boldsymbol{p}_g^* - C_g(\boldsymbol{p}_g^*, \boldsymbol{x}_g^*, \boldsymbol{u}_g^*)), \tag{3.3}$$

Since the ISO's decision $(\boldsymbol{p}^*, \boldsymbol{x}^*, \boldsymbol{u}^*)$ is a feasible, but not necessarily optimal solution to problem (3.1)–(3.2), $U_g(\boldsymbol{\pi}, \boldsymbol{p}^*, \boldsymbol{x}^*, \boldsymbol{u}^*)$ is non-negative.

The *convex hull prices* are defined to be the dual maximizers $\boldsymbol{\pi}^*$ of the Lagrangian dual of UC (2.14). The value function of the Lagrangian dual problem (2.14) as a function of $\boldsymbol{d}$ is the convex envelope of $v(\boldsymbol{d})$ [23]. The price vector is a sub-gradient of the convex envelope of $v(\boldsymbol{d})$.

The duality gap between the UC problem and its dual (2.13) is exactly

the total lost opportunity costs,

$$\sum_{g \in \mathcal{G}} U_g(\boldsymbol{\pi}, \boldsymbol{p}^*, \boldsymbol{x}^*, \boldsymbol{u}^*). \tag{3.4}$$

Consequently, convex hull pricing as a uniform pricing scheme minimizes total uplift payment as defined by (3.4); that is, it minimizes the total lost opportunity costs of all participating units. This is a special case of a more general result on the type of uplift payments convex hull pricing minimizes, as shown in [10, 29, 75], since the only type of system-wide constraint considered for now is the supply-demand balance constraint.

3.3 Primal Formulation

Theorems 2.3–2.6 imply that, if each unit faces only generation limits and minimum up/down constraints, exact convex hull prices can be determined by solving **LD-Primal** with the convex hulls explicitly described in Section 2.4.2 and convex envelopes explicitly described in Section 2.4.3. The convex hull prices are the optimal dual variables associated with the supply-demand balance constraints (2.16). When ramping constraints are present, Section 2.5.3 gives a close yet tractable approximation.

Since the number of constraints in our explicit formulation of **LD-Primal** is polynomial in T and $|\mathcal{G}|$, and since all variables are continuous, the convex hull pricing problem can be solved as a convex program in polynomial time. Note that in an optimal solution to **LD-Primal**, the commitment and start-up variables can be fractional. In convex hull pricing, we focus on the

optimality of dual variables. The ISO's commitment and dispatch decisions are still determined by the UC problem.

When there are transmission constraints or other system-wide constraints, as discussed in Section 2.5.1, these constraints are unaltered in **LD-Primal**. The locational prices can be derived as a function of the dual variables associated with the supply-demand balance constraints and the transmission constraints, as in locational marginal pricing [87, Section 8.11]. Note that in the presence of system-wide constraints that do not necessarily hold as equalities at a welfare-maximizing solution, such as the transmission constraints, the gap between the UC problem and its dual includes not only the total lost opportunity cost of the units, but also another type of uplift that addresses the ISO's revenue insufficiency [10, 75].

We next consider two market design issues for convex hull pricing in non-convex electricity markets: when uplift is only paid to some market participants, and when prices are calculated on a rolling horizon basis.

3.3.1 Minimization of Uplift Payments to a Subset of the Participating Units

In certain electricity markets, only a subset of the participating units $\mathcal{G}$ can receive uplift payments. For example, it may be the case that only units dispatched to a strictly positive generation level are qualified for uplift payments; that is, units are not paid for merely participating in the market. In this case, the duality gap that convex hull pricing minimizes includes terms

that might not end up being paid as uplift.

Let the set of units that are qualified to receive uplift payments be $\mathcal{G}' \subseteq \mathcal{G}$. If the qualifications can be determined prior to computing the prices (this is the case where $\mathcal{G}'$ is the set of units actually committed by the ISO), we can solve **LD-Primal** with $\mathcal{G}$ replaced by $\mathcal{G}'$. The duality gap that the resulting prices minimize includes only uplift payments to units in $\mathcal{G}'$.

3.3.2 Committing to Prices

In convex hull pricing, prices that are coupled across multiple time periods as a whole minimize uplift payments over the specified horizon T of the underlying UC problem. A subset of these prices does not necessarily minimize uplift payments over any shorter time horizon that is a subset of T. In a day-ahead market, the whole set of 24 hourly prices is calculated and posted at once, so that the coupling would not be problematic insofar as the ISO commits to buying and selling at these prices.

In contrast, look-ahead real-time markets are operated on a rolling basis where only the commitment, dispatch, and price calculated for the upcoming interval are implemented. Therefore, the coupling inherent across a *single* look-ahead dispatch may not be represented appropriately by the sequence of convex hull prices, each of which corresponds to the upcoming interval in each successive look-ahead dispatch.

To make prices consistent across successive look-ahead dispatches, [35] suggests that the pricing model represent past intervals in the convex hull

pricing model, keep the commitment and dispatch decisions in the past as variables, but constrain prices in the past intervals to be equal to the realized prices. A simple way to achieve this in **LD-Primal** is to dualize the system-wide constraints corresponding to each realized price with a penalty equal to the realized price. For example, if no transmission constraint is binding in a previous period, we dualize the supply-demand balance constraint corresponding to that period with a penalty equal to the realized price. If there have been binding transmission constraints in previous periods, we can dualize each of those binding transmission constraints with a penalty equal to its realized optimal dual variable.

3.4 Numerical Results

We implement **LD-Primal** on a personal computer with a 2.2-GHz quad-core CPU and 16 GB of RAM. The optimization problems are modeled in CVX [28] and solved with GUROBI 6.5 [31]. We consider four examples from the literature. The time resolution in all examples is one hour.

3.4.1 Example 1

We consider an example from [75] in which two units (including a block-loaded one) serve 35 MW of load in a single period. We modify the original example by including a start-up cost for each unit. Table 3.1 specifies each unit's offers. Both units are assumed to be off initially. The optimal (and the only feasible) solution to the ISO's UC problem is for unit 1 to generate 35

MW and for unit 2 to stay offline.

Table 3.1: Supply Offers in Example 1

Unit	Start-up	No-load	Energy	$\underline{p}_g$	$\overline{p}_g$
	\$	\$	\$/MWh	MW	MW
1	100	0	50	10	50
2	100	0	10	50	50

Table 3.2: Comparison of Different Pricing Schemes for Example 1

Pricing Scheme	π	U_1	U_2
	\$/MWh	\$	\$
LMP	50	100	1900
CHP	12	1430	0
CHPq	52	30	-

Unit 1 is marginal[4] and sets the LMP. Seeing the LMP, unit 2's profit-maximizing decision is to go online and generate 50 MW, which would result in a profit of \$1900. Therefore, based on the definition of U_g in (3.3), unit 2 has a lost opportunity cost of \$1900. The start-up cost of unit 1 is not covered by LMP. An uplift payment of \$100 is needed to make unit 1 whole (guarantee a non-negative profit). This result is summarized in the LMP row of Table 3.2.

[4]A marginal unit has an optimal dispatch level strictly between its maximum and minimum power output.

LD-Primal for this example is:[5]

$$\min_{p_g, x_g, u_g, g=1,2} \quad 100u_1 + 50p_1 + 100u_2 + 10p_2 \tag{3.5}$$

$$\text{s.t.} \quad p_1 + p_2 = 35 \tag{3.6}$$

$$10x_1 \leq p_1 \leq 50x_1 \tag{3.7}$$

$$50x_2 \leq p_2 \leq 50x_2 \tag{3.8}$$

$$u_1 \geq x_1, \quad u_2 \geq x_2 \tag{3.9}$$

$$u_1 \geq 0, \quad u_2 \geq 0. \tag{3.10}$$

Theorems 2.3 and 2.4 imply that the exact convex hull price can be obtained as the optimal dual variable associated with the demand constraint. Table 3.2 shows that the resulting uplift payments are lower than those under LMP but still quite large.

Suppose that only units dispatched to a strictly positive generation level are qualified for uplift payments. In this situation, unit 2 does not receive any compensation for its lost opportunity cost. As suggested in Section 3.3.1, to minimize the uplift payment to the qualified units, we can instead solve **LD-Primal** with unit 2 excluded. We refer to this pricing scheme as CHPq (CHP for qualified units). Table 3.2 shows that, if only unit 1 is qualified for compensation, CHPq results in a lower uplift payment ($30) than CHP ($1430).

[5]Since the cost functions in this example are linear, our primal formulation is a linear program.

3.4.2 Example 2

We investigate a three-period two-unit example from [51] with ramping constraints but without startup costs. Table 3.3 shows the supply offers and ramp rate limits. For this example, the start-up/shut-down ramp rate limit $\overline{v}_g$ of each unit equals the unit's ramp-up/down rate when committed, v_g. All units are assumed to be off initially. Table 3.4 presents the optimal commitment and dispatch decisions as well as the demand in each period. Ramping constraints require unit 2 to commit at $t = 2$ so that it can ramp up to the generation level needed at $t = 3$. Table 3.5 displays energy prices and uplift payments under different pricing schemes.

Since unit 1 is the marginal unit in all three periods, the LMPs are set by unit 1 at \$60/MWh. The payment based on LMPs covers all of unit 1 costs. As shown in the first row of Table 3.5, an uplift payment of \$560 is needed to "make unit 2 whole".

We can approximate $\text{conv}(\mathcal{X}_g)$ with constraints (2.18)–(2.21), (2.24), and ramping constraints. We refer to this pricing method as aCHP1 (approximate CHP). In aCHP2, we augment our formulation with valid inequalities describing $\text{conv}(\mathcal{X}_g)$ with $T = 2$. Finally, using the description of $\text{conv}(\mathcal{X}_g)$ for $T = 3$, we formulate the convex hulls in this three-period example, which results in the exact convex hull prices.

Table 3.5 shows the energy prices and uplift payments under different pricing schemes. For this example, as the approximation of $\text{conv}(\mathcal{X}_g)$ becomes

Table 3.3: Units in Example 2

Unit	No-load	Energy	$\underline{p}_g$	$\overline{p}_g$	Ramp Rate
	\$	\$/MWh	MW	MW	MW/hr
1	0	60	0	100	120
2	600	56	0	100	60

Table 3.4: Optimal Commitment and Dispatch for Example 2

t	d_t	$x_{1,t}$	$p_{1,t}$	$x_{2,t}$	$p_{2,t}$
	MW		MW		MW
1	70	1	70	0	0
2	100	1	40	1	60
3	170	1	70	1	100

Table 3.5: Comparison of Different Pricing Schemes for Example 2

Pricing Scheme	π_1	π_2	π_3	U_1	U_2
	\$/MWh	\$/MWh	\$/MWh	\$	\$
LMP	60	60	60	0	560
aCHP1	60	60	64	120	160
aCHP2	60	60	65.6	168	0
CHP	60	60	65.6	168	0

more accurate, the energy price at $t = 3$ increases. Roughly speaking, unit 1 has an increasing incentive to generate more than the ISO's optimal dispatch, increasing its lost opportunity cost, but unit 2's no-load costs can be better covered, decreasing its lost opportunity cost. The net effect is a decrease in total uplift as π_3 increases.

Note that the prices resulting from aCHP2 happen to equal the exact

convex hull prices. This result implies that the valid inequalities for $T = 3$ are "non-binding". The approximation of $\text{conv}(\mathcal{X}_g)$ in aCHP2 is accurate enough to yield the exact convex hull prices for this example.

3.4.3 Example 3

We consider a 24-period 32-unit example from [83]. The cost functions for the units are linear. There are no ramping or transmission constraints. **LD-Primal** is an LP through which the exact convex hull prices can be obtained. **LD-Primal** solves in 0.02 seconds, resulting in the same convex hull prices as reported in [83]. The duality gap between the UC problem and its Lagrangian dual problem is \$1 148, which equals the total lost opportunity cost.

3.4.4 Example 4

We consider a 96-period 76-unit 8-bus example that is based on structural attributes and data from ISO New England [40]. We consider the start-up costs, no-load costs, minimum up/down time constraints, and ramping constraints for the generation units. The cost functions are quadratic. We use Scenario 1 of the 90 load scenarios provided in [40]. Minimum generation levels for the units are not specified in the original data, so we let $\underline{p} = 0.8\overline{p}$ for each nuclear plant and $\underline{p} = 0.6\overline{p}$ for each coal-fired unit. The units' initial statuses are not provided. We solve a single-period UC problem to obtain the optimal commitment and dispatch decisions for period 1. We use these optimal decisions as the units' initial statuses, and assume that the units have been on/off

for sufficiently long time so that the minimum up/down time constraints are not initially binding. The flow limits of the 12 transmission lines in the system are not defined in the original data. Therefore, we first investigate the case without transmission constraints (Case 1). We then set a limit of 2100 MW on the flow over each transmission line (Case 2).

For each case, we first solve the UC problem and obtain the LMPs. We then determine convex hull prices using two methods: an approximation of **LD-Primal** and the single-period approximation proposed in [82]. Because of the ramping constraints, we approximate $\text{conv}(\mathcal{X}_g)$ using constraints (2.18)–(2.21), (2.24), along with valid inequalities that completely characterize these convex hulls for $T = 2$ and $T = 3$. We use the convex envelope described in Theorem 2.5 and solve the primal formulation as an SOCP. When solving the UC problems, we include above-mentioned valid inequalities a priori, and set MIPgap to 0.01%.

Table 3.6 shows the results for the UC problem and the approximated **LD-Primal**, as well as the relative gap between these two problems. The approximated **LD-Primal** solves in polynomial time with respect to the number of constraints. However, if we were to solve the Lagrangian dual problem in the dual space for Case 2, there would be 2400 dual variables. Such a large number of dual variables due to transmission constraints creates difficulties for non-smooth optimization methods [85].

The relative gap between approximated **LD-Primal** and the UC problem (called CHP gap hereafter) is 0.06% for Case 1 and 0.10% for Case 2.

Table 3.6: UC and approximated LD-Primal for Example 4

	Case 1	Case 2
UC Obj.($)	41 972 688	42 303 772
UC CPU Time (s)	87.42	112.02
LD-Primal Obj. ($)	41 946 298	42 259 962
LD-Primal CPU Time (s)	6.22	13.04
Gap (%)	0.063	0.104

This small CHP gap bounds two other gaps from above. First, the duality gap between the UC problem and its Lagrangian dual problem can only be smaller than the CHP gap. This verifies the theoretical result shown in [7], which states that the relative duality gap of the UC problem and its Lagrangian dual approaches zero as the number of heterogeneous generators approaches infinity. Second, the approximation error (the gap between the conceptual **LD-Primal** and our approximation) is bounded from above by the CHP gap.

Table 3.7 compares the total uplift payment under the three pricing schemes. We only consider units' lost opportunity costs. In the single-period approximation, start-up and no-load costs are considered only for fast-start units. We classify a unit with a minimum up/down time of one hour as a fast-start unit, and 18 units fall into this category. We allocate start-up costs to peak usage hours according to the recommendation in [82]. In both cases, each single-period approximation solves in much less than a second. The convex hull prices derived from the proposed method result in the least uplift payment in both cases, with and without transmission constraints.

Table 3.7: Total Lost Opportunity Cost ($) for Example 4

	LMP	Primal Formulation for CHP	Single-Period Approximation of CHP
Case 1	183 473	33 965	96 938
Case 2	329 032	40 863	177 391

Chapter 4

Generation Expansion Planning with Consideration of Unit Commitment

4.1 Introduction

Large shares of renewable generation in electric power systems increase the need for operational flexibility, which is limited by ramping constraints and minimum up/down time constraints. Consideration of operational flexibility in GEP necessitates the modeling of UC in system operations. However, the problem of UC itself is computationally difficult. Therefore, to incorporate operational flexibility, researchers have made efforts in improving two main approaches of GEP, screening curve methods and mathematical optimization.

Screening curve methods have been widely adopted in GEP because of its computational efficiency [6, 66, 91]. In these methods, the net demand[1] is modeled through a load duration curve (LDC), an aggregation of the demand time series. This aggregation does not preserve the chronological information

This chapter is based on the following publication: Bowen Hua, Ross Baldick, and Jianhui Wang. "Representing Operational Flexibility in Generation Expansion Planning Through Convex Relaxation of Unit Commitment." *IEEE Transactions on Power Systems* 33.2 (2018): 2272-2281. The first author designed the model, designed and implemented the computational studies, and wrote the manuscript with support from coauthors.

[1]The net demand is defined as electric demand minus uncontrollable renewable generation.

that captures temporal variation of renewable generation. As a result, temporal constraints of generating units (e.g., minimum up/down time constraints and ramping constraints) cannot be fully represented in the planning problem. Such loss of temporal information poses as a major difficulty for representing limits on operational flexibility. To approximately consider flexibility-related issues, [6, 91] use a heuristic short-term operational model that makes unit commitment decisions.

In the mathematical optimization approach, GEP is typically modeled as a mixed-integer program (MIP). Binary decision variables represent investment decisions, and an embedded system operational model makes short-term operational decisions; total investment and operational costs are minimized. Ideally, to fully represent operational flexibility, the embedded operational model needs to be a unit commitment (UC) problem that represents chronological demand and renewable generation. In this ideal situation, we explicitly model discrete commitment decisions and temporal variations of load and renewable generation, thereby enabling the consideration of key operational flexibility issues (e.g., binary commitment decisions, minimum up/down time, ramping constraints). Annual operational costs can also be accurately determined. However, the computational complexity of the UC problem prevents this ideal. A few previous studies propose methods for incorporating unit commitment constraints in an optimization-based GEP model. In these studies, various types of approximations are introduced to reduce the computational burden, including using a representative load curve and/or simplifying the

embedded operational model.

For instance, Ma *et al.* [48] consider a GEP problem with a UC-based operational model. To manage the computational burden, they select four representative weeks, in which the chronological load and renewable generation curves are carefully determined to represent a whole year. In order for simulated operational cost to be accurate, the load and renewable generation curves in a representative week need to be a certain type of average among the represented weeks. In particular, the representative weeks need to capture the "worst-case" variation of net load, as opposed to a simple average. These contradictory objectives, together with the need to consider the correlation of variation between demand and renewable generation, make the determination of the representative load and renewable generation curves a difficult task.

Other studies have adopted simplifications of the embedded UC model in GEP. For example, ramping and operating reserves are considered in [19] within a linear programming formulation. The commitment of the individual units and minimum up/down time constraints are not considered. It is reported in [56] that the commercial software PLEXOS uses a linear programming relaxation of the UC model to manage computational complexity. Also, PLEXOS approximates the hourly chronological load curve by fitting the curve with blocks that are coarser in temporal resolution.

A promising unit clustering approach has been proposed to reduce computational complexity in the embedded operational model [60–63]. In this method, units within the same cluster are assumed to have identical techno-

logical parameters (capacity, heat rate, ramp rate, etc.), i.e., the heterogeneity of units within the same cluster is assumed away. Instead of using a binary commitment vector for each unit, an integer commitment vector is used for each cluster of units. The clustering strategy is crucial in controlling the error introduced by such approximation. Although not explicitly stated by the authors of [60–63], finding an optimal clustering strategy that minimizes clustering error is itself a difficult combinatorial optimization problem. Thus, several heuristic clustering strategies are proposed in [62]. Moreover, unit clustering has only limited value if transmission constraints are considered and similar units are not co-located.

In this chapter, following the line of research that simplifies the operational UC model, we adopt **LD-Primal** proposed in Chapter 2 as the embedded operational model. The Lagrangian dual problem provides a tight relaxation of UC, as the relative duality gap between the UC problem and its Lagrangian dual approaches zero as the number of heterogeneous generators approaches infinity [7,8]. Compared to the standard bi-level formulation of the Lagrangian dual problem, **LD-Primal** is single-level, thus can be embedded into a GEP problem.

Section 4.2 presents the GEP model previously published in [36] that adopts **LD-Primal** as the embedded operational model. Our model is computationally efficient, because the embedded operational model is a continuous and polynomially-solvable optimization problem; the only integer variables in our GEP model are the investment decisions. Section 4.3 demonstrates the

computational efficiency of our GEP model and the tightness of our convex relaxation using a Texas system in which we consider a chronological load curve of 8760 hours.

4.2 GEP Formulation

To incorporate operational flexibility in GEP in a computationally efficient manner, we propose a MIP formulation of the GEP problem that employs **LD-Primal** as the operational model.

4.2.1 Investment Constraints

Let $\mathcal{G}_{\mathrm{N}}$ denote the set of candidate new generating units, $\mathcal{G}_{\mathrm{E}}$ the set of existing units, and $\mathcal{G} = \mathcal{G}_{\mathrm{N}} \cup \mathcal{G}_{\mathrm{E}}$.

The only integer variables in our GEP model are the binary investment variables z_g. The variable z_g equals one when candidate unit $g \in \mathcal{G}_{\mathrm{N}}$ is built and zero otherwise. The following constraints take investment decisions into consideration:

$$x_{gt} \leq z_g, \forall t \in [0, T], \forall g \in \mathcal{G}_{\mathrm{N}}. \tag{4.1}$$

4.2.2 Private Constraints

We introduce private constraints that apply to each $g \in \mathcal{G}$. We consider the co-optimization of energy and spinning reserve. We consider minimum up/down time constraints and ramping constraints. We include valid

inequalities that characterize the convex hull of an individual unit's feasible region when $T = 2$. Since the models in Section 2.5.3 do not consider spinning reserve, we introduce the following constraints for all $t \in [2, T]$:

$$p_{gt} - p_{g,t-1} + r_{gt} \leq v_g x_{g,t-1} + \overline{v}_g(1 - x_{g,t-1}), \tag{4.2}$$
$$p_{g,t-1} - p_{gt} + r_{gt} \leq v_g x_{gt} + \overline{v}_g(1 - x_{gt}), \tag{4.3}$$
$$p_{g,t-1} + r_{g,t-1} \leq \overline{v}_g x_{g,t-1} + (\overline{p}_g - \overline{v}_g)(x_{gt} - u_{gt}), \tag{4.4}$$
$$p_{gt} + r_{gt} \leq \overline{p}_g x_{gt} - (\overline{p}_g - \overline{v}_g)u_{gt}, \tag{4.5}$$
$$p_{gt} - p_{g,t-1} + r_{gt} \leq (\underline{p}_g + v_g)x_{gt} - \underline{p}_g x_{g,t-1} - (\underline{p}_g + v_g - \overline{v}_g)u_{gt}, \tag{4.6}$$
$$p_{g,t-1} - p_{gt} + r_{gt} \leq \overline{v}_g x_{g,t-1} - (\overline{v}_g - v_g)x_{gt} - (\underline{p}_g + v_g - \overline{v}_g)u_{gt}. \tag{4.7}$$

Using results in Sections 2.4.2, 2.5.3, and 2.5.2, for each unit $g \in \mathcal{G}$ in our GEP formulation, the feasible region is defined to be:

$$\{\boldsymbol{p}_g \in \mathbb{R}^T, \boldsymbol{x}_g \in \mathbb{R}^T, \boldsymbol{u}_g \in \mathbb{R}^{T-1}, \boldsymbol{r}_g \in \mathbb{R}^T \mid$$
$$(2.18)\text{–}(2.21), (2.24), (2.35), (2.36), (4.2)\text{–}(4.7)\},$$

which is a tractable approximation of $\text{conv}(\mathcal{X}_g)$. We note that, although we only consider spinning reserve, our formulation can be extended to incorporate other types of reserve.

4.2.3 System-Wide Constraints

We include the following system-wide constraints:

- power balance constraints (2.11),
- reserve requirement constraints:

$$\sum_{g \in \mathcal{G}} \boldsymbol{r}_g \geq \underline{\boldsymbol{r}}, \tag{4.8}$$

where $\underline{\boldsymbol{r}} \in \mathbb{R}_+^T$ is a constant vector for system-wide spinning reserve requirement determined by operational protocols of the system operator.

We note that any other linear system-wide constraints can be incorporated without affecting the tightness of **LD-Primal**, since the system-wide constraints are not dualized in the Lagrangian relaxation.

4.2.4 Objective Function

Since we concern ourself with system operations during a year, we consider annualized capital cost c_g^{cap}, an amortization of the overnight capital cost (OCC) of generation investment over the lifespan (LS) of the unit [79]:

$$c_g^{\text{cap}} = \frac{\gamma \times OCC}{1 - 1/(1+\gamma)^{LS}}, \tag{4.9}$$

where γ is the interest rate. Let parameter c_g^{FOM} denote the annual fixed operations and maintenance (O&M) costs.

The objective function is:

$$C^{\text{total}} = \sum_{g \in \mathcal{G}_{\text{N}}} z_g (c_g^{\text{cap}} + c_g^{\text{FOM}}) + \sum_{g \in \mathcal{G}_{\text{E}}} c_g^{\text{FOM}} + \sum_{g \in \mathcal{G}} C_g^{\text{var},**}. \tag{4.10}$$

Recall that we use **LD-Primal** as the embedded operational model. Therefore, instead of using the variable cost function, we include the convex envelope of the variable cost function in the objective function.

The explicit formulation of the convex envelopes depends on the form of the variable cost function, which may include startup costs, no-load costs, fuel costs, and variable O&M costs. We consider three cases:

1. The variable cost function is affine in $\boldsymbol{p}_g$. In this case we have

$$C_g^{\text{var}} = \sum_{t=1}^{T} b_g p_{gt} + c_g x_{gt} + h_g u_{gt}, \tag{4.11}$$

where h_g denotes the startup cost, c_g the no-load cost, and b_g a constant. Since this function is affine in its arguments, its convex envelope has the same functional form.

2. The variable cost function is piecewise linear in $\boldsymbol{p}_g$. Let $\underline{p}_g$ and $\overline{p}_g$ be the minimum and maximum generation levels for unit g, respectively. Partition the interval $[\underline{p}_g, \overline{p}_g]$ into $|\mathcal{K}|$ intervals:

$$[\underline{p}_g, \overline{p}_g] = \bigcup_{k \in \mathcal{K}} \mathcal{I}_k, \tag{4.12}$$

where k is the index for the partitioned intervals. Let h_g be the startup cost, and let c_g denote the no-load cost. Suppose at each period t, when $p_{gt} \in \mathcal{I}_k$, the operating cost (excluding start-up cost and no-load cost) is

$$\tilde{C}_{gt}(p_{gt}, 1, 0) = a_{gk} p_{gt} + b_{gk}. \tag{4.13}$$

We have the variable cost function in this case:

$$C_g^{\text{var}} = \sum_{t=1}^{T} \Big\{ a_{gk} p_{gt} + b_{gk} + c_g x_{gt} + h_g u_{gt}, \text{ if } p_{gt} \in \mathcal{I}_k \Big\}. \tag{4.14}$$

According to the results in Section 2.4.3, its convex envelope is also a convex piecewise linear function:

$$C_g^{\text{var},**} = \sum_{t=1}^{T} \Big\{ a_{gk} p_{gt} + (c_g + b_{gk}) x_{gt} + h_g u_{gt}, \text{ if } \frac{p_{gt}}{x_{gt}} \in \mathcal{I}_k \Big\}. \tag{4.15}$$

To implement the convex envelope, we can introduce a new variable $s_{gt} \in \mathbb{R}$ for each $g \in \mathcal{G}$ and $t \in [0, T]$, define

$$C_g^{\text{var},**} = \sum_{t=1}^{T} s_{gt} + c_g x_{gt} + h_g u_{gt}, \tag{4.16}$$

and include the following constraints

$$s_{gt} \geq b_{gk} p_{gt} + b_{gk} x_{gt}, \forall k \in \mathcal{I}_k. \tag{4.17}$$

The resulting GEP is a mixed-integer linear program.

3. The variable cost function is quadratic in $\boldsymbol{p}_g$. We have

$$C_g^{\text{var}} = a_g p_{gt}^2 + b_g p_{gt} + c_g x_{gt} + h_g u_{gt}, \tag{4.18}$$

where $a_g > 0$ and b_g are constants. Constant h_g denotes the startup cost and c_g denotes the no-load cost. According to the results in Section 2.4.3, the convex envelope of the variable cost function is:

$$C_g^{\text{var},**} = \sum_{t=1}^{T} \begin{cases} a_g \frac{p_{gt}^2}{x_{gt}} + b_g p_{gt} + c_g x_{gt} + h_g u_{gt}, & x_{gt} > 0, \\ 0, & x_{gt} = 0. \end{cases} \tag{4.19}$$

4.2.5 Formulations of GEP

We first present our GEP model that represents limits on operational flexibility through convex relaxation of UC, **GEP-UC-R**:

$$\begin{aligned} \min_{\Delta_{\text{UC}}} \quad & C^{\text{total}} && && (4.20) \\ \text{s.t.} \quad & (2.18)\text{–}(2.20), (2.24), (2.35), (2.36), (4.2)\text{–}(4.7), && \forall g \in \mathcal{G} && (4.21) \\ & (4.1), (2.11), (4.8), && && (4.22) \end{aligned}$$

where C^{total} is specified in (4.10), and

$$\Delta_{\text{UC}} = \{z_g, g \in \mathcal{G}_{\text{N}}, \tag{4.23}$$

$$\boldsymbol{p}_g, \boldsymbol{x}_g, \boldsymbol{u}_g, \boldsymbol{r}_g, \, g \in \mathcal{G}\}. \tag{4.24}$$

To investigate the importance of considering operational flexibility in GEP, we consider another formulation of GEP in which constraints on operational flexibility are ignored, **GEP-ED**:

$$\min_{\Delta_{\text{ED}}} \quad \sum_{g \in \mathcal{G}_{\text{N}}} z_g (c_g^{\text{cap}} + c_g^{\text{FOM}}) + \sum_{g \in \mathcal{G}_{\text{E}}} c_g^{\text{FOM}} + \sum_{g \in \mathcal{G}} C_g^{\text{var,ED}} \tag{4.25}$$

$$\text{s.t.} \quad 0 \leq p_{gt} \leq z_g \overline{p}_g, \quad \forall t \in [1, T], \quad \forall g \in \mathcal{G}_{\text{N}} \tag{4.26}$$

$$0 \leq p_{gt} \leq \overline{p}_g, \quad \forall t \in [1, T], \quad \forall g \in \mathcal{G}_{\text{E}} \tag{4.27}$$

$$(2.11), \tag{4.28}$$

where $C_g^{\text{var,ED}}$ is a variable cost function that excludes start-up cost, and

$$\Delta_{\text{ED}} = \{z_g, g \in \mathcal{G}_{\text{N}}, \tag{4.29}$$

$$\boldsymbol{p}_g, g \in \mathcal{G}\}. \tag{4.30}$$

4.3 Numerical Results

We conduct case studies on a test case based on the Electric Reliability Council of Texas (ERCOT) system. By comparing the generation expansion plans obtained from different GEP models, we show that it is necessary to consider limits on operational flexibility in GEP. We demonstrate that **LD-Primal** provides an accurate yet tractable model for system operations, enabling the representation of 8760-hour chronological system operations, a task that would have been computationally formidable without the convex relaxation.

4.3.1 Test System

We adopt technical and cost parameters of the existing generating units in ERCOT (as of year 2014), predicted chronological electricity generation from renewables for the year 2030, and predicted chronological load for the year 2030 reported in [21]. To fully consider the hourly, weekly, and seasonal variability of load and renewable generation, we use hourly chronological load and renewable generation predictions for a whole year (8760 hours), avoiding the error that would have been introduced by selecting representative time periods or by reducing temporal resolution.

We exclude the following sets of existing units in our planning problem: 1) units that would have retired in the year 2030; 2) hydro units, which amount to a very small fraction (less than 1%) of the generation capacity in ERCOT; 3) units with uncommon fuel types, such as biomass. The existing thermal generation mix (a total of 150 units) consists of: 19.67 GW of coal-fired units, 5.02 GW of simple-cycle gas-fired units, 35.20 GW of gas-fired combined-cycle units, and 5.13 GW of nuclear units. The peak predicted system demand in the year 2030 is 81.21 GW, and the mean system demand is 57.14 GW.

We consider the aggressive renewable scenario in [21], which predicts 33 846 MW of installed wind capacity, 3 091 MW of installed solar generation, and provides a sample path of chronological renewable generation in the year 2030. We model investments in renewable generation as an input, rather than decision variables, since such investments are in part policy-driven. The operations and maintenance costs of renewables are not included in the total

cost. In the embedded operational model, we allow economic curtailment of renewables without penalty, and treat the dispatch of renewables, up to available production dictated by the forecast, as continuous decision variables.

For new generation investments, we consider the following types of candidate thermal units:

- Coal: coal-fired steam turbines,
- CCGT: combined cycle gas turbines,
- SCGT: simple cycle gas combustion turbines,
- Nuclear: nuclear units.

Table 4.1 shows input data for candidate generation units.

Table 4.1: Parameters of the Candidate Generating Units

	Coal	CCGT	SCGT	Nuclear
Unit Size (MW)	650	410	200	1215
ACC[2] (10^3\$/MW·yr)	174.7	60.1	38.5	224.0
Heat Rate (MMBTU/MWh)	8.00	7.34	14.31	13.62
Variable O&M (\$/MWh)	7.33	4.73	13.41	1.33
Fixed O&M (10^3\$/MW·yr)	43.56	25.28	16.95	58.62
Startup Cost (\$/MW·start)	54.11	16.23	28.14	100
Min. Up Time (hr)	24	6	1	48
Min. Down Time (hr)	12	12	1	24
Ramp Rate (p.u.)	0.3	0.5	1	0.1
Min. Generation (p.u.)	0.5	0.3	0.25	0.8

[2]ACC = Annualized Capital Cost.

We use the reference case of the year 2030 in the Energy Information Administration's (EIA's) projection [81] for fuel costs. EIA does not predict cost of uranium, and data from [71] is used to estimate the cost of uranium. The fuel costs are converted to 2014 dollar and shown in Table 4.2.

Table 4.2: Fuel Costs

	$/MMBTU
Coal	2.89
Natural Gas	5.78
Uranium	0.87

We consider an affine variable cost function (constant marginal cost) for each unit. As a result, the embedded operational problem LD-Primal is a linear program (LP), and GEP-UC-R is a mixed-integer linear program. We do not model the transmission system. We do not consider outages of the units. The spinning reserve requirement is set to be the maximum of the following two values: 1) total capacities of two largest generators, 2) the sum of 3.3% of load and 15% of wind power forecast. In all the GEP models, we model the reserve requirement constraints as soft constraints, and penalize reserve shortage at $1 000/MWh, a penalty ISO New England uses for 30-minute operating reserve [1]. We model the power balance constraints as hard constraints.

We implement our model on a personal computer with a 2.2-GHz quad-core CPU and 16 GB of RAM in MATLAB. The optimization problems are modeled in CVX [28] and solved with GUROBI 6.5 [31]. When solving the

planning models, to reduce memory usage, we choose the dual simplex method to solve the integer relaxations in the branch and bound process, and set the thread count to one. We set the convergence criterion as a MIP gap of 0.01%.

4.3.2 Optimal Expansion Plans

We compare the optimal expansion plans obtained from different GEP formulations. We name each optimal expansion plan after the model that is used to obtain the plan.

Section 4.2.5 introduced two GEP formulations, **GEP-UC-R** and **GEP-ED**. In addition, we implement a GEP model based on the unit clustering approach proposed in [63].[3] We cluster the existing thermal units by technology and heat rate, which results in one cluster for nuclear, three clusters for coal, two clusters for SCGT, and three clusters for CCGT. The capacity of a single unit in each cluster is determined by the average among all units in that cluster. The technological parameters of a single unit in each cluster is determined through an average weighted by the true capacity of each unit. Candidate units of the same technology are assumed to be identical. Therefore, we cluster the candidate units by technology only (coal, CCGT, SCGT, nuclear). We refer to the clustered UC model as UC-C. We refer to the GEP model that embeds UC-C as GEP-UC-C.

[3]Apart from different assumptions such as outages and ancillary services, we make two modifications in our implementation compared to [63]. First, due to what we believe is a typo in the original paper, we modify the first item in equation (19) of [63] from 1 to N_g (notation in [63] used here). Second, we model the reserve constraints as soft constraints in GEP, whereas [63] models them as hard constraints.

Table 4.3: Computational Time and Objective Function Values

	CPU Time (s)	Optimal Obj. Value ($)
GEP-UC-R	97371	1.286×10^{10}
GEP-UC-C	82921	1.288×10^{10}
GEP-ED	749	1.275×10^{10}

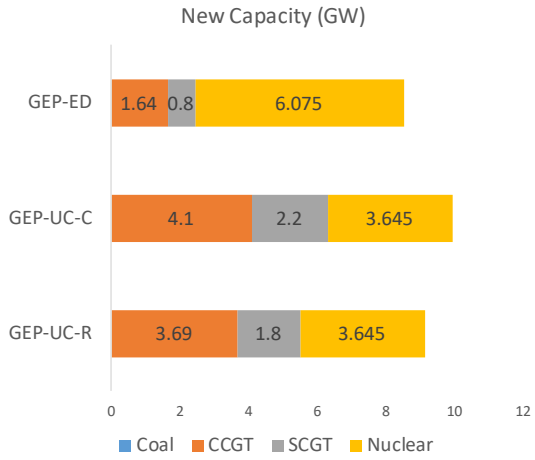


Figure 4.1: Comparison of optimal generation expansion plans.

Fig. 4.1 shows the differences between the optimal expansion plans. None of the plans involves any investment in coal. GEP-ED invests in more nuclear generation, while GEP-UC-R and GEP-UC-C include more flexible gas-fired units, presumably because the representation of limits on operational flexibility. The total capacity investments obtained from GEP-ED (8.515 GW) is less than that from GEP-UC-R (9.135 GW), which is less than that from GEP-UC-C (10.145 GW).

Table 4.3 shows computational time and optimal objective function value of the three GEP models. Despite appearing to have a lower optimal

objective function value, GEP-ED fails to incorporate commitment-related issues, which leads to a suboptimal expansion plan and even infeasibility in system operations, as we show in Section 4.3.3.

The difference in optimal objective values of GEP-UC-R and GEP-UC-C can be explained by different expansion plans (more gas-fired units in GEP-UC-C) and different approximation errors. We examine the latter issue in Section 4.3.4.

4.3.3 System Operations

To better illustrate the optimality of the expansion plans and evaluate the impacts of neglecting operational flexibility, we use a full UC model to simulate system operations in the year 2030. In this simulation, we take generation investment decisions as input.

Since solving an ERCOT-size UC problem of 8760 hours is computationally impractical, we simulate system operations by solving a series of UC problems on a rolling basis, similar to how day-ahead markets operate in the US. The units' operational status at the ending period of one UC problem serves as the initial status of the subsequent UC problem. Each unit commitment problem concerns 48 hours of system operations, except for the last one ($8760 = 48 \times 181 + 72$). We choose the length of the simulation to be equal to the longest minimum up/down time of the units. We refer to such yearly operations simulation as Ops-UC.

In the operational UC problems, we model the power balance con-

straints as hard constraints, and the reserve requirement constraints as soft constraints, which is consistent with the GEP models. However, as we will see later, under the GEP-ED expansion plan, UC problems for certain time periods would be infeasible if we keep the power balance constraints as hard constraints. In this case, we re-solve the UC problem with the power balance constraints converted to soft constraints with a penalty of \$9 000/MWh, the value of lost load used by ERCOT [22]. The rolling simulation then continues.

Table 4.4: System Operations in the Year 2030

			GEP-UC-R	GEP-UC-C	GEP-ED
Energy (TWh)	Existing	Coal	64.18	64.19	56.08
		CCGT	122.74	122.78	130.17
		SCGT	3.93	3.79	4.17
		Nuclear	44.96	44.97	44.96
	New	Coal	0.00	0.00	0.00
		CCGT	25.86	25.90	10.21
		SCGT	1.07	1.11	0.47
		Nuclear	31.93	31.93	53.13
	Renewables		59.02	59.02	58.96
RC (GWh)			67.90	63.77	124.42
LS (GWh)			0.00	0.00	2.23
RS (GW·h)			66.94	32.28	94.22
OC ($\$\times 10^9$)			11.91	11.89	11.49
IC ($\$\times 10^9$)			1.16	1.19	1.64
TC ($\$\times 10^9$)			13.07	13.08	13.14

*RC = Renewable Curtailment, LS = Load Shedding, RS = Reserve Shortage, OC = Operational Cost, IC = Investment Cost, TC = Total Cost

Table 4.4 shows the results for system operations under the three generation expansion plans, including energy generation from different technologies, renewable curtailment, load shedding, reserve shortage, and costs. We note that, under GEP-ED, the operational UC problem is infeasible for periods

110, 115, and 117 (which are all summer weeks) if the power balance constraints are modeled as hard constraints. In these three periods, we re-solve the operational UC problem by penalizing the violation of power balance constraints. Since GEP-ED does not take commitment issues into consideration, not enough flexible units are built, so that the power balance constraints cannot be satisfied during the periods with dramatic net load variation. Such lack of operational flexibility in the GEP-ED plan results in a total of 2.2 GWh load shedding. In contrast, power balance constraints for all hours are satisfied under GEP-UC-R and GEP-UC-C.

Since the GEP-ED expansion plan invests in more nuclear generation, more energy is generated from nuclear, so that the total operational cost is lower than that of the other two plans. However, such low cost is achieved at the price of more renewable curtailment, reserve shortage, and load shedding. Also, the capacity factor of the existing SCGT units is higher under the GEP-ED plan, because these units have to be utilized and ramped more often due to lack of operational flexibility in the GEP-ED plan. Moreover, the investment cost of the GEP-ED plan is higher due to the high construction cost of nuclear units. In sum, the total annual cost under GEP-ED is the highest of the three expansion plans.

When compared to GEP-UC-R, GEP-UC-C invests in more gas-fired units, leading to less reserve shortage and less operational cost. However, GEP-UC-R results in the least total annual cost of the three plans.

We note that, under the GEP-UC-R plan, the total annual operational

cost obtained in Ops-UC ($\$1.191 \times 10^{10}$) is higher than the total operational costs evaluated in the planning model ($\$1.170 \times 10^{10}$). The same pattern is also observed for the other two expansion plans. This may be attributed to: 1) suboptimality of solving UC problems on a rolling basis as compared to solving a single UC problem of 8760 hours; 2) relaxation gap of **LD-Primal**. We investigate the relaxation gap in the next subsection.

4.3.4 Computational Efficiency and Tightness of Convex Relaxation

In this subsection, we solve the three operational problems (full UC, our convex relaxation **LD-Primal**, and the unit-clustering model UC-C) to illustrate the tightness and computational efficiency of **LD-Primal**. We fix the expansion plan to be GEP-UC-R, the plan that leads to the least annual cost out of the three expansion plans.

Similar to Ops-UC in which a series of UC problems are solved, we simulate yearly system operations by solving a series of **LD-Primal** and UC-C on a rolling basis. The initial status of the units for each **LD-Primal**/UC-C problem is set to be identical with the initial status of the corresponding time period in Ops-UC. We refer to such operations simulations as Ops-UC-R and Ops-UC-C, respectively.[4]

[4]We believe that reserve constraints were modeled as hard constraints in [63] when power balance constraints are relaxed for feasibility, because no information on reserve shortage is given even when there is load shedding. In our implementation, we model the reserve constraints as soft constraints whose penalty is lower than the value of lost load.

When solving the operational models, we set the thread count to eight. We set the convergence criterion for the full UC problems and UC-C problems as a MIPgap of 0.01%. The optimization problems solved in Ops-UC-R are continuous. Although both **LD-Primal** and UC-C have the potential to solve a large-scale problem for 8760 hours, we choose the rolling-basis approach in the operations simulation because the benchmark problem, the full UC model, is computationally too difficult to be solved as a monolithic problem of 8760 hours.

Table 4.5 shows the computational results of the three operations simulation models under the GEP-UC-R plan. The absolute error is defined to be the gap in the summed optimal objective value (annual operational cost) between Ops-UC and its simplified model. While Ops-UC-C requires less computational time, Ops-UC-R clearly outperforms Ops-UC-C in terms of accuracy. This result shows that, for the system that we consider, **LD-Primal** is a better approximation of the UC problem compared with the unit clustering approach.

Although we do not consider the transmission system, since system-wide constraints are dualized in the Lagrangian relaxation, **LD-Primal** could be easily extended to incorporate transmission constraints. In contrast, in a transmission-constrained system, unit clustering that does not take location into consideration would introduce more error.

We have previously observed that the annual operational cost obtained in Ops-UC is 1.76% higher than the annual operational cost evaluated in the

Table 4.5: System Operations Under GEP-UC-R Plan

	Ops-UC	Ops-UC-R	Ops-UC-C
CPU Time (s)	7094	883	511
OC ($)	1.1914×10^{10}	1.1909×10^{10}	1.1939×10^{10}
Absolute Error ($)	-	4.978×10^{6}	2.502×10^{7}
Error (%)	-	0.042%	0.21%

planning model. Since the relaxation gap of **LD-Primal** is only 0.042%, we can conclude that the higher operational cost can be largely attributed to the suboptimality of solving UC on a rolling basis.

Part III

Modeling of Combined Cycle Units

Chapter 5

Tight Formulation of Transition Ramping of Combined Cycle Units

5.1 Introduction

A combined cycle unit (CCU) is a type of generator that consists of one or more combustion turbines (CTs), each with a heat recovery steam generator (HRSG), as well as one or more steam turbines (STs) (see Fig. 5.1 for example). Steam produced by each HRSG is used to drive STs.

Based on different combinations of CTs and STs, a CCU can be operated in one of several *configurations*. This is different from the formulation in Chapter 2 where we assume binary on/off states for each generation unit. Certain rules have to be followed for the transitions between different configurations; not all transitions are feasible. For instance, to switch on an ST, at least one CT must be on. Fig. 5.2 shows an example of feasible transitions for a CCU with two CTs and one ST, known as a 2-on-1 CCU.

As described in the introduction, large shares of renewable generation in electric power systems increase the need for operational flexibility. As a result, many combined cycle units (CCUs) have been built as a major provider of flexibility in power systems. Whereas in the past CCUs might run for

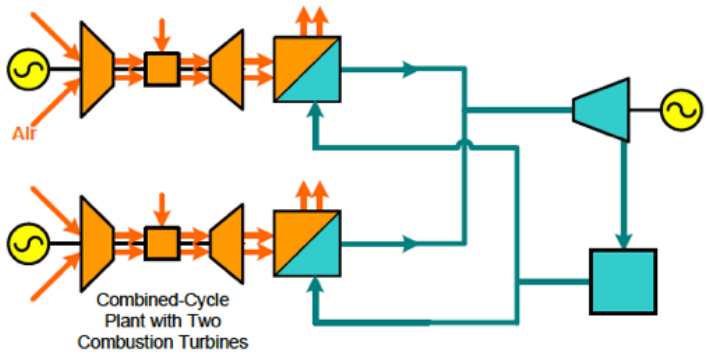


Figure 5.1: Typical configurations of a 2-on-1 (2CT + 1ST) CCU. The orange arrows indicate gas flows, and the cyan arrows indicate steam flows. The orange trapezoids represent CTs, the cyan trapezoids represent STs, the cyan squares represent condensers, and the bi-color squares represent HRSGs. Source: [3].

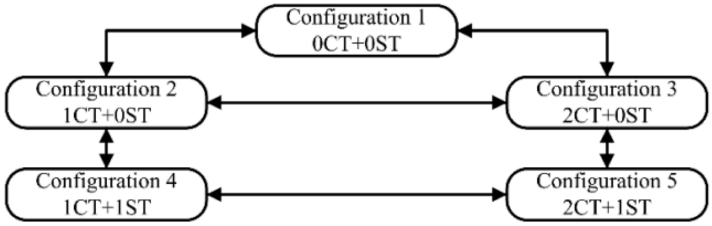


Figure 5.2: Configurations and feasible transitions of a 2-on-1 CCU. Source: [52].

lengthy periods of time serving base load, they now respond quickly to variations in renewable supply and demand by frequent ramping, as well as more frequent changing of configurations (transitions). In addition to determining the configuration of CCUs in day-ahead unit commitment, optimizing their configurations in real time (e.g., through the look-ahead commitment) can pre-position CCUs to cope with the variability of renewables.

Because of the flexible, cyclic pattern of use, the modeling of the transitions becomes increasingly important. As we will show in Section 5.2, most

existing modeling approaches of CCUs assume that units are considered to start/end their production within one interval. This assumption can often be violated, especially in look-ahead commitment and dispatch models where the length of the interval is typically fifteen minutes or even five minutes. To accurately model the transitions, we propose a mixed-integer programming model for the transition ramping of CCUs in Section 5.3. We also show the tightness of the proposed model using our polyhedral results from Chapter 2. Section 5.4 presents a numerical example.

5.2 Existing Modeling Approaches

5.2.1 Overview of Existing Approaches

Currently, many ISOs in the US (e.g., MISO, NYISO, PJM, and ISO-NE) use an aggregated modeling approach for CCUs in their UC model [24]. The aggregated modeling approach assumes that at each time interval each CCU may either be on or off, which is only a rough approximation for CCUs that have multiple configurations. As the number of CCUs increases, ISOs such as ERCOT, CAISO, and SPP have moved toward more detailed modeling of CCUs.

Besides the aggregated approach, there are mainly two types of modeling approaches of CCUs in the unit commitment problem. The first approach is the component-based (or physical-unit-based) modeling [15]: each of the physical units of a CCU is represented by a set of commitment and dispatch variables. This approach has been recognized as more suitable for security

analysis than for market-clearing [38], in part because many technical parameters would have to be submitted to the ISO had this approach been adopted.

Another approach is the configuration-based modeling in which each configuration of a CCU is represented by a set of commitment and dispatch variables [12, 44]. This approach is viewed as suitable for market clearing and bid/offer processing [2, 38], and is adopted by CAISO and ERCOT [52]. However, as pointed out by [24], the standard configuration-based approach cannot describe the minimum up/down time constraints of each individual CT and ST in a CCU. To this end, an edge-based formulation is proposed in [24] in which the minimum run time constraints of each individual turbine are captured, at the cost of introducing more variables into the formulation. In addition, a new configuration-based model is proposed in [17] where the minimum run time constraints of individual turbines are formulated in the configuration-based variables via projection.

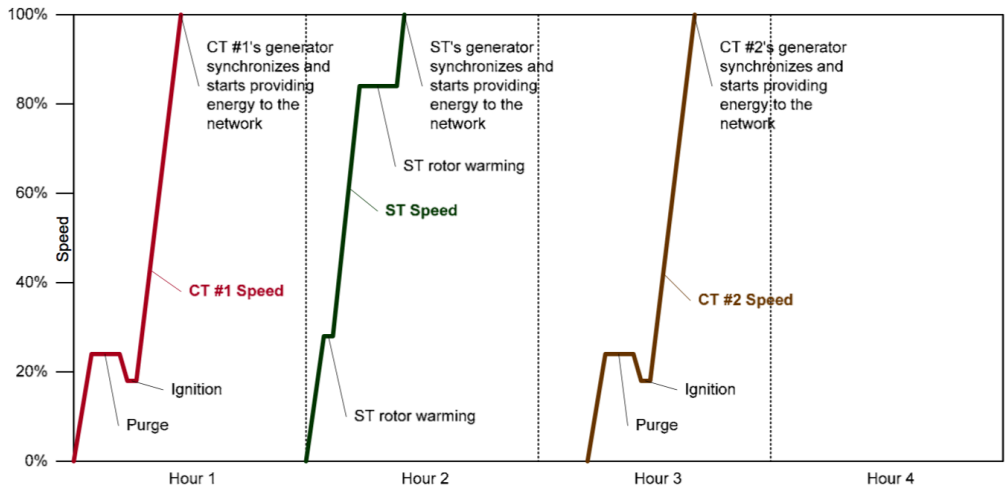
There have been efforts in obtaining tight and compact formulations for UC, a task closely related to the characterization of the convex hull of an individual unit's feasible region and the convex envelope of its cost function (Chapter 2). The tightness of a MIP formulation refers to how close the MIP's integer relaxation is to the MIP itself. The compactness of a MIP formulation refers to its size. Using the minimum up/down time constraints proposed by [69], [59] show that the three-binary (on/off, startup, shutdown) formulation of UC leads to better computational performance than the one-binary (on/off) formulation [11] mainly due to tightness. The three-binary

formulation is further improved in [54]. A tight and compact formulation for UC with configuration-based modeling of CCUs is presented in [52].

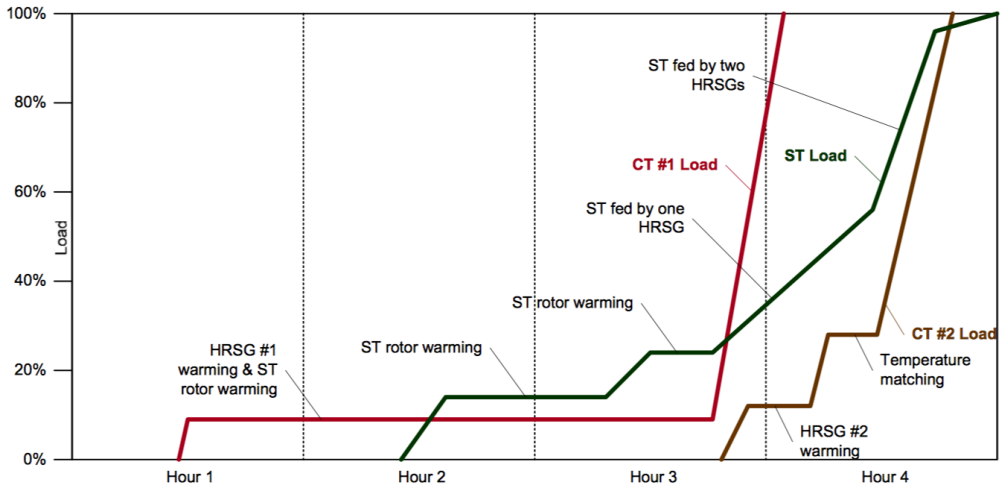
5.2.2 Issues with the Existing Models for CCU Transition

For physical reasons, combined-cycle units have little ability to follow an exterior control signal during start-up (synchronization and ramping up to minimum output) and shut-down process, although the unit injects power into the grid after synchronization. The CCU's electrical output is reasonably predictable during the start-up process. The output during start-up, while fairly predictable, does not increase smoothly from minimum load to maximum load. Instead, output during start-up is characterized by extended holds, where plant output does not change, periods where plant output increases slowly, and periods where plant output increases rapidly [3]. As we can see from Fig. 5.3, during the start-up of this CCU, the operations of the two CTs and one ST are coordinated.

In addition to the starting up of the entire CCU, turbines of the CCU might also start up during the transitions between different configurations. The total power output of a CCU during transitions can usually be modeled as a fixed trajectory. For example, for a 2-on-1 (two CTs and one ST) CCU to transition from 2CT to 2CT+1ST, typically the CTs have to reach maximum output before the ST starts up. In the meantime, the output from ST during its start-up is also predictable. As a result, the total power output from the CCU (maximum output of the two CTs plus the start-up power-trajectory of



Component Speeds – Representative Cold Start



Component Loadings – Representative Cold Start

Figure 5.3: Cold start-up of a typical CCU with two CTs and one ST. Source: [3]

the ST) during the transition can be viewed as a known trajectory.

The start-up and shut-down trajectories of simple-cycle units have been studied in [53]. However, most existing modeling approaches for CCUs do not reflect the above-mentioned transition behavior: they assume that units start up, shut down, and transition[1] within one interval while the transition ramping is ignored. This assumption has the following problems:

- the transition times are usually longer than the time length of a single interval, which invalidates the assumption. The short interval length of look-ahead commitment and dispatch (typically 15 minutes) exacerbates this problem.
- CCUs are always modeled as dispatchable. As a result, the actual CCU output during transition might be different from the ISO's dispatch in the day-ahead or look-ahead commitment problem. The difference has to be compensated by real-time dispatch or regulation.
- Moreover, it has been observed in practical real-time markets that CCUs during transition may submit a ramp rate limit of 0.1 MW as a proxy of the non-dispatchable status of the unit.

To sum up, most existing CCU models make the invalid assumption that all transitions are completed within a single interval. This assumption leads to a discrepancy between the model and the reality, as well as suboptimal commitment and dispatch. In this chapter, we propose a mixed-integer

[1] We will use transition as an unified term for start-up, shut-down, and transition between different configurations hereafter.

programming model for CCUs where we remove the invalid assumption and model the power output of CCUs in transition as a fixed trajectory.

5.3 Mathematical Formulation

In this section, we describe the mathematical formulation of the proposed CCU model with representation of transition ramping. For brevity, we only show the constraints that define the feasible region of a single CCU. Embedding these equations into a complete MIP formulation of unit commitment is straightforward.

Specifically, we first introduce a standard configuration-based formulation for CCU where we assume the completion of any transition within a single interval. Next, we describe a new formulation where we remove the aforementioned assumption, and show the tightness of the proposed formulation.

5.3.1 Standard Configuration-Based Formulation

We show a standard configuration-based formulation from [52]. Let y index the set of configurations $\mathcal{Y}$. Let $t \in \{1, \ldots, T\}$ be the index for the time interval.

Decision Variables. In configuration-based modeling, we have a set of binary variables x for each of the configurations, and a set of transition variables v for each of the feasible transitions. The decision variables are:

x_t^y Binary variable for whether configuration $y \in \mathcal{Y}$ is *on* at t;

$v_t^{y,y'}$ Binary variable for transition from $y \in \mathcal{Y}$ to $y' \in \mathcal{Y}$ at t;

p_t^y Continuous variable for power output from configuration $y \in \mathcal{Y}$ at t;

p_t Continuous variable for power output of the CCU at t.

Constraints.

- Bounds on the power output of each configuration:

$$\underline{p}^y x_t^y \leq p_t^y \leq \overline{p}^y x_t^y, \, \forall t, \forall y \in \mathcal{Y}, \tag{5.1}$$

  where $\underline{p}^y$ and $\overline{p}^y$ are respectively the lower and upper bound of power output from configuration y.

- Total power output of the CCU:

$$p_t = \sum_{y \in \mathcal{Y}} p_t^y, \, \forall t. \tag{5.2}$$

- Configurations are mutually exclusive:

$$\sum_{y \in \mathcal{Y}} x_t^y = 1, \, \forall t. \tag{5.3}$$

- Logical relationship between configuration and transition variables:

$$x_t^y - x_{t-1}^y = \sum_{y' \in \mathcal{M}^{T,y}} v_t^{y'y} - \sum_{y' \in \mathcal{M}^{F,y}} v_t^{yy'}, \, \forall t, \forall y \in \mathcal{Y}. \tag{5.4}$$

  where $\mathcal{M}^{F,y}$ is the set of reachable configurations from $y \in \mathcal{Y}$, and $\mathcal{M}^{T,y}$ is the set of configurations that can reach $y \in \mathcal{Y}$.

- At most one transition in each interval:

$$\sum_{yy' \in \mathcal{M}} v_t^{yy'} \leq 1, \forall t, \tag{5.5}$$

  where $\mathcal{M}$ is the set of all feasible transitions.

We note that there might be additional constraints characterizing the feasible set of a CCU, including ramping and minimum run time of each configuration or turbine. However, since these constraints are irrelevant to the representation of the transition between configurations, we omit them here. The feasible set of a CCU under the single-interval transition assumption is defined by constraints (5.1)–(5.5).

Next, we extend this standard configuration-based formulation by removing the invalid assumption of single-interval transition.

5.3.2 Proposed Formulation

Additional Data for Transition Modeling. Recall that we model the total power output from CCU as a fixed trajectory because CCU has little ability to follow an external control signal during transition. To represent the power-trajectories for each feasible transition, we introduce additional data that describe the length of the transition and the shape of the trajectory:

- Let $TP_i^{yy'}$ be the total power output from the CCU in transition at the end of the i-th interval of the transition process between y and y'.

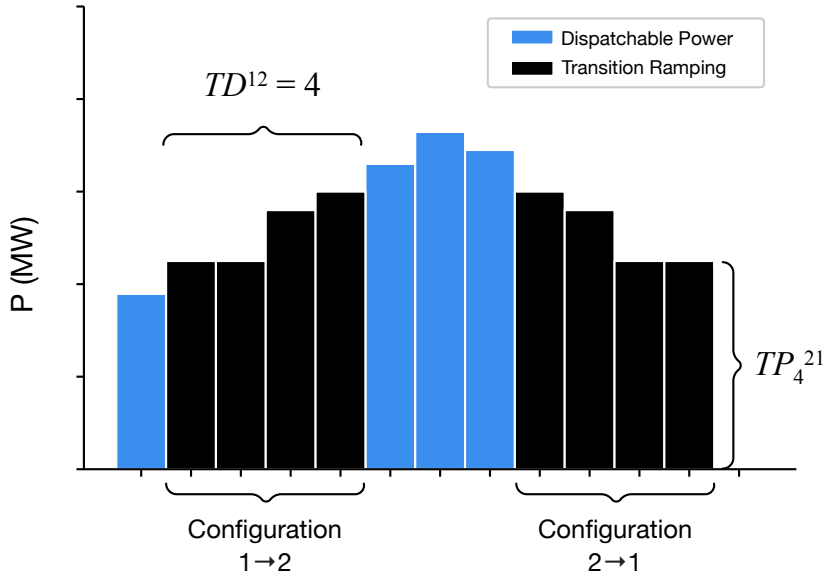


Figure 5.4: Illustration of additional data for transition ramping. The horizontal axis shows the interval and the vertical axis shows the power output of the CCU at each interval.

- Let $TD^{yy'}$ be the duration (number of intervals) of the transition process between y and y'.

The vector $[TP_1^{yy'}, TP_2^{yy'}, \ldots, TP_{TD^{yy'}}^{yy'}]$ describes the transition power-trajectory.

Fig. 5.4 illustrates the two sets of additional data we need for modeling transition ramping. The CCU shown in the figure operates in configuration 1 at the beginning, transitions to configuration 2, and then transitions back to configuration 1. The duration of the transition from configuration 1 to 2, as well as 2 to 1, is $TD^{12} = TD^{21} = 4$ intervals. The power output during the four-interval transition from configuration 2 to 1 is specified by the 4-vector $(TP_1^{21}, TP_2^{21}, TP_3^{21}, TP_4^{21})$, with TP_4^{21} shown explicitly in Fig. 5.4.

Decision Variables. The decision variables of the proposed formulation are:

w_t^y Binary variable for whether configuration y is *dispatchable* at t;

x_t^y Binary variable for whether configuration y is either 1) dispatchable or 2) on the from end of an ongoing transition at t;

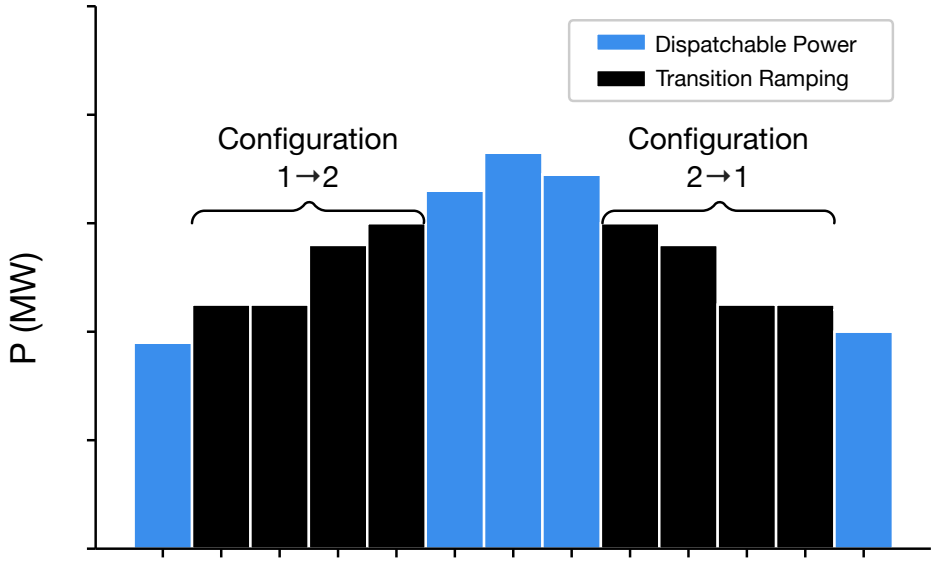
$v_t^{y,y'}$ Binary variable for transition from y to y' at t. The transition variable becomes one when a new configuration *becomes dispatchable*;

p_t^y Continuous variable for power output from configuration $y \in \mathcal{Y}$ at t;

p_t Continuous variable for power output of the CCU at t.

We note that we introduce a new set of binary variables $\boldsymbol{w}$ for dispatchability. Fig. 5.5 illustrates this new variable using the same transition events shown in Fig. 5.4. At the first interval, configuration 1 is dispatchable, and $w_1^1 = 1$. The transition from configuration 1 to 2 lasts four intervals, and w_t^2 does not become 1 until interval 6. Similarly, when transition back to configuration 1, w_t^1 does not become one until interval 13.

Compared to the standard formulation, the definition of p_t and p_t^y remains the same, although p_t^y does not include the output implied by the transition power-trajectories. To adjust for the transition intervals when the CCU is non-dispatchable, $v_t^{y,y'}$ is defined to be one when a new configuration becomes



t	1	2	3	4	5	6	7	8	9	10	11	12	13
x^1	1	1	1	1	1	0	0	0	0	0	0	0	1
x^2	0	0	0	0	0	1	1	1	1	1	1	1	0
w^1	1	0	0	0	0	0	0	0	0	0	0	0	1
w^2	0	0	0	0	0	1	1	1	0	0	0	0	0

Figure 5.5: Illustration of the new binary variable w. The horizontal axis shows the time interval and the vertical axis shows the power output of the CCU at each interval. The table below shows the values of the variables x_t^y and w_t^y for $t \in [1, 13]$ and $y \in \{1, 2\}$.

dispatchable. That is, the transition variable $v_t^{y,y'}$ is one when the transition completes. The definition of x_t^y is changed accordingly so that constraints (5.4) still hold.

Constraints. We keep constraints (5.3)–(5.5) from the standard formulation. In addition, we have:

- Bounds on the power output of each configuration:

$$\underline{p}^y w_t^y \leq p_t^y \leq \overline{p}^y w_t^y, \, \forall t, \forall y. \tag{5.6}$$

- Constraint that defines w_t^y:

$$w_t^y = x_t^y - \sum_{yy' \in \mathcal{M}} \sum_{i=1}^{TD^{yy'}} v^{yy'}_{t-i+1+TD^{yy'}}, \, \forall t, \forall y. \tag{5.7}$$

  - The last term forces w_t^y to zero when CCU is transitioning away from configuration y.

- Total power output of CCU:

$$p_t = \sum_{y \in \mathcal{Y}} p_t^y + \sum_{yy' \in \mathcal{M}} \sum_{i=1}^{TD^{yy'}} TP_i^{yy'} v^{yy'}_{t-i+1+TD^{yy'}}, \, \forall t. \tag{5.8}$$

  - The last term represents the output from the transition power-trajectory.

Again, here we omit ramping and minimum up/down time of configuration/turbine. The feasible set of a CCU with representation of transition ramping is defined by constraints (5.3)–(5.8).

5.3.3 Analysis of the Proposed Formulation

5.3.3.1 Tightness of the Proposed Formulation

To study the tightness of the proposed formulation, we focus on the following feasible set:

$$\mathcal{S} = \{(\boldsymbol{x} \in \{0,1\}^{T|\mathcal{Y}|}, \boldsymbol{v} \in \{0,1\}^{T|\mathcal{M}|}, \boldsymbol{w} \in \{0,1\}^{T|\mathcal{Y}|}, \boldsymbol{p} \in \mathbb{R}^{T|\mathcal{Y}|}) | (5.3)\text{–}(5.7)\}, \tag{5.9}$$

where $\boldsymbol{x}$ is a vector that collects all x_t^y variables, $\boldsymbol{v}$ collects all $v_t^{y,y'}$ variables, $\boldsymbol{w}$ collects all w_t^y variables, and $\boldsymbol{p}$ collects all p_t^y variables. Note that we omit p_t variables in our polyhedral study. This is because 1) they are defined through equality constraints (5.8), and thus can be eliminated; 2) they do not affect the tightness of the formulation, as shown by the polyhedral result from Chapter 2.

Remark 5.1. Without ramping constraints, if we can describe the convex hull of the feasible set of the binary variables:

$$\mathcal{S}_{\text{bin}} = \{(\boldsymbol{x} \in \{0,1\}^{T|\mathcal{Y}|}, \boldsymbol{v} \in \{0,1\}^{T|\mathcal{M}|}, \boldsymbol{w} \in \{0,1\}^{T|\mathcal{Y}|}) | (5.3)\text{–}(5.5), (5.7)\}, \tag{5.10}$$

then by Theorem 2.2 applied to $\mathcal{S}$, the convex hull description together with constraints (5.6) define the convex hull of the whole feasible set $\mathcal{S}$.

The constraints that link the binary and the continuous variables are (5.6). For Theorem 2.2 to hold, the double-sided inequality constraints (5.6) have to be of the form of

$$B(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{w}) \leq A\boldsymbol{p} \leq D(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{w}), \tag{5.11}$$

and satisfy:

1. The number of rows of A has to be less than the number of continuous variables p_t^y.
2. Each entry of $D - B$ is nonnegative.

When constraints (5.6) are expressed in the form of (5.11), we see that $A = I$, $B = \mathbf{0}$, and D is a diagonal matrix with nonnegative entries. In addition, the number of rows of A equals the length of the p vector. Therefore, Theorem 2.2 holds.

Remark 5.1 implies that, once we determine the convex hull of the binary variables, characterizing the convex hull of the whole feasible set is trivial. However, characterizing the convex hull of the binary variables is itself a difficult problem, especially in the presence of minimum run time constraints (see a discussion in [52]).

5.3.3.2 Compactness of the Proposed Formulation

Since variables w_t^y are defined through equality constraints (5.7), we can replace all appearances of w_t^y in the other constraints by the right hand side of (5.7), thereby reducing the number of variables as well as the number of constraints. Same applies to variables p_t defined through equality constraints (5.8).

Remark 5.2. If we eliminate all w_t^y and p_t variables in the proposed formulation using (5.7), then the number of constraints and the number of variables in the proposed formulation remain the same as those in the standard formulation.

If we eliminate all w_t^y and p_t variables, then the set of decision variables $(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{p})$ is the same as the standard formulation. The number of constraints does not change, either, because constraints (5.6) and (5.8) replace (5.1) and

(5.2), respectively. Compared with the standard formulation, the proposed formulation only makes the coefficient matrix denser.

5.3.3.3 Additional Aspects

Ramping Constraints. As with the single-cycle units, the introduction of ramping constraints complicates the convex hull of the CCU feasible set $\mathcal{S}$ and invalidates Remark 5.1.

In the existing literature such as [52], both intra-configuration and inter-configuration ramp rates are defined. However, the inter-configuration ramp rate is only a rough proxy to the transition power-trajectories.

We propose to use plant-wise ramping constraints:

$$p_t - p_{t-1} \leq \sum_{y \in \mathcal{Y}} R_U^y x_t^y, \ \forall t, \tag{5.12}$$

$$p_{t-1} - p_t \leq \sum_{y \in \mathcal{Y}} R_D^y x_t^y, \ \forall t, \tag{5.13}$$

where R_U^y and R_D^y are respectively the ramp-up and ramp-down rate limits of configuration y when the CCU is dispatchable.

For some CCUs, the CT has to reach its maximum output before committing the ST. Plant-wise ramping constraints take such transition rules into consideration. In addition, plant-wise ramp rate leads to fewer constraints compared to the existing formulations.

Constraints (5.12) and (5.13) assume that the ramp rate of the power-trajectories are within the limit on the ramp rates when the CCU is dispatch-

able. If this assumption is not satisfied, we can introduce additional terms on the right hand side that relax these constraints when the CCU is in transition.

No Overlapping Transitions. Since we remove the one-interval transition assumption in the proposed formulation, one may worry about overlapping transitions. That is, the start of a new transition before the completion of a previous transition.

The minimum run time constraints may prevent such overlapping. More importantly, we show that overlapping is not possible in the proposed formulation even without the minimum run time constraints.

Remark 5.3. The proposed formulation (5.3)–(5.8) does not allow overlapping transitions.

To see this, suppose there are two overlapping transitions, first from y to y', then from y' to y''. The durations these two transitions are $TD^{y,y'}$ and $TD^{y',y''}$, respectively. Without loss of generality, we assume that the first transition starts at interval 1. Since the two transitions are overlapping, the starting time s of the second transition has to satisfy $1 < s \leq TD^{y,y'}$.

Because the first transition has not completed at s, $x_s^y = 1$, $w_s^y = 0$, and $x_s^{y'} = 0$. At $t = s$, constraint (5.7) specifies:

$$w_s^{y'} = x_s^{y'} - \sum_{i=1}^{TD^{y'y''}} v_{t-i+1+TD^{y'y''}}^{y'y''}.$$

Since the second transition starts at s, the second term on the right hand side

is one. The above constraint implies $w_s^{y'} = -1$, which contradicts $w_s^{y'} \in \{0, 1\}$. Therefore, there can be no overlapping transitions.

5.4 Numerical Results

To illustrate the proposed CCU model, we use an example where we solve a multi-interval unit commitment problem.

5.4.1 Problem Data

Consider a look-ahead commitment and dispatch problem with one simple-cycle unit, one CCU, 8 intervals, and no transmission constraints. The demand vector is $\boldsymbol{d} = [370, 425, 435, 450, 480, 520, 560, 580]$.

The simple cycle unit has a minimum output of 0 MW, a maximum output of 200 MW, a constant marginal cost of \$5/MWh, and a start-up cost of \$200. Ramp rate limits are not considered.

The combined cycle unit has two CT, one ST, and a duct burner (DB)[2]. Table 5.1 shows the five configurations of the CCU.

The feasible transitions of the CCU, the transition costs, and transition power-trajectories are shown in Table 5.2. If the length of a transition is zero, then we assume that transition is completed within a single interval because we do not have data for the power-trajectory of that transition.

[2]Duct burners increase the steam production from heat recovery steam generators (HRSGs) to above what is possible with just the heat from the CT exhaust.

Table 5.1: Configurations of the Combined Cycle Unit

Configuration	Marginal Cost	No-load Cost	$\underline{p}^y$	$\overline{p}^y$
	\$	\$/MWh	MW	MW
1: All off	0	0	0	0
2: 1CT+1ST	18.99	1596	150	225
3: 2CT+1ST	17.98	4457	330	445
4: 1CT+1ST+DB	26.33	1596	225	250
5: 2CT+1ST+DB	23.94	4457	445	525

Table 5.2: Feasible Transitions of the Combined Cycle Unit

From Config	To Config	Transition Cost	Length	Trajectory
		\$	Intervals	MW
1	2	16854	0	
1	3	49486	0	
2	3	36807	4	$[225, 240, 280, 300]$
2	4	12000	2	$[225, 225]$
2	1	450	0	
3	2	36450	4	$[300, 280, 240, 225]$
3	5	20000	2	$[445, 445]$
3	1	900	0	
4	2	10000	2	$[225, 225]$
5	3	18000	2	$[445, 445]$

Initially, the simple cycle unit generates 200 MW of power; the CCU is in configuration 2, with 150 MW of power output. We ignore the ramping constraints and the minimum run time constraints.

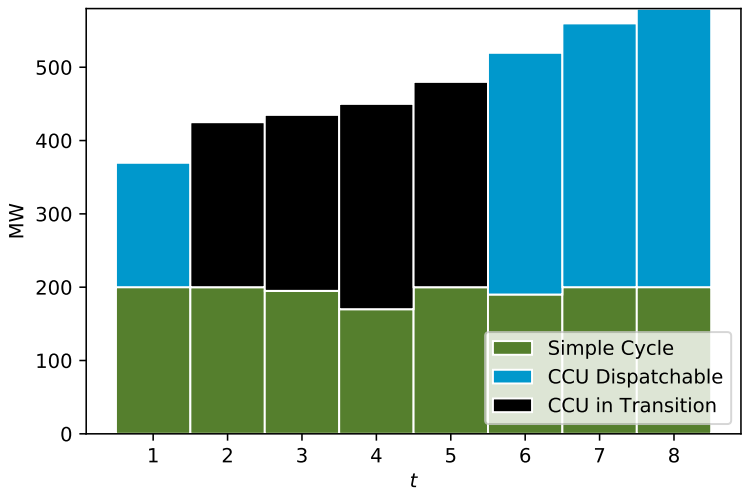


Figure 5.6: Optimal dispatch with the proposed CCU model. The horizontal axis indicates the time interval, and the vertical axis represents the MW output. The height of each stacked bar is the system-wide load. Each segment of a stacked bar indicates the portion of load served by different resources.

5.4.2 Results

Fig. 5.6 shows the optimal commitment and dispatch solved with the proposed model. To satisfy the increasing demand, the CCU starts to transition from configuration 2 (1CT + 1ST) to configuration 3 (2CT + 1ST) at interval 2, bringing online the second combustion turbine. The transition lasts four intervals, during which the total power output from CCU is a pre-defined trajectory, shown as the dark green bars in Fig. 5.6. Starting from interval 6, the CCU becomes dispatchable again.

In comparison, Fig. 5.7 shows the optimal commitment and dispatch solved with the standard model where we make the invalid assumption of

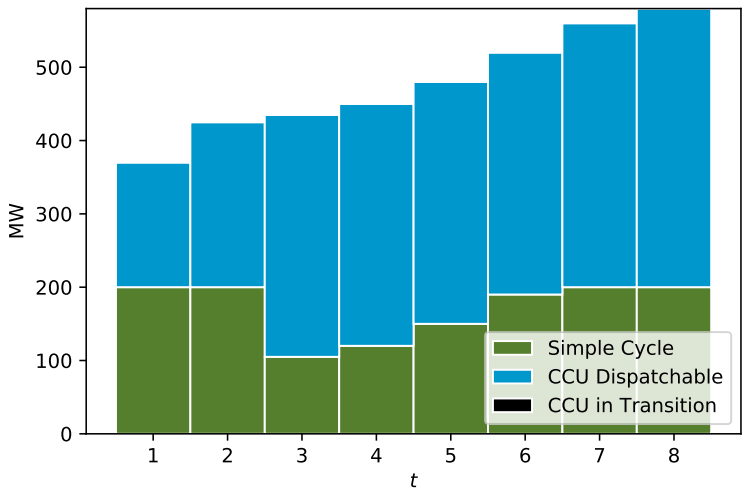


Figure 5.7: Optimal dispatch with the standard CCU model. The horizontal axis indicates the time interval, and the vertical axis represents the MW output. The height of each stacked bar is the system-wide load. Each segment of a stacked bar indicates the portion of load served by different resources.

single-interval transitions. The CCU transitions from configuration 2 to configuration 3 at interval 3, bringing online another CT. However, due to the invalid assumption we make, the CCU becomes instantly dispatchable at interval 3, generating 330 MW of power.

The power outputs of the CCU during the first four intervals after the transition are $[330, 330, 330, 330]$ MW, which deviates from the transition power-trajectory $[225, 240, 280, 300]$ MW. Because the CCU has limited ability to follow exterior control signal during the transition, this discrepancy makes the dispatch and commitment solutions sub-optimal.

Part IV

Pricing in Multi-Interval Real-Time Markets

Chapter 6

Pricing in Multi-Interval Real-Time Markets

6.1 Introduction

Large shares of renewable generation in electric power systems have increased the variability of net load. To better manage the variabilities, several wholesale markets in the US have been contemplating or have implemented a *multi-interval real-time market* (MIRTM). Instead of solving a single-interval economic dispatch problem, MIRTM is based on a *look-ahead dispatch* (LAD)[1] problem that considers several intervals. By expanding the time horizon, MIRTM allows for more efficient dispatch of generation to meet projected system conditions by pre-positioning resources to cope, for example, with large ramps in net load [32, 88, 89]. In particular, a satisfactory short-term forecast for renewable generation can allow LAD to meet high ramping requirements due to renewable generator variability in a reliable and economic manner. Consequently, the New York Independent System Operator (NYISO) and California ISO (CAISO) have already implemented MIRTM, while

[1]Some MIRTMs might involve commitment decisions, so that the underlying optimization problem is a look-ahead commitment and dispatch problem. In this chapter, we focus on the pricing of the dispatch problem, as the pricing issues related to the non-convex commitment decisions have been addressed in Chapter 3.

the Electric Reliability Council of Texas (ERCOT) has proposed the approach (see [68], [50], [90], and [25]).

In a typical implementation of MIRTM, the ISO solves LAD in a fashion that resembles model predictive control (MPC) [89]. Fig. 6.1 shows such an implementation. For each dispatch problem, only the dispatch decisions for the *upcoming* interval[2] are used; dispatch decisions for the later intervals are *advisory*. Similarly, in a typical implementation of MIRTM, prices for the upcoming interval are used for settlement; we refer to such prices as *settlement prices* hereafter. Prices for intervals further in the future may be published but are not financially binding; we refer to these prices as *advisory prices* hereafter.

If the ISO did not need to solve dispatch problems on a rolling horizon basis, then the optimal price sequence from a single MIRTM problem would support the ISO's multi-interval dispatch instructions.[3] However, in a MPC-like implementation of MIRTM, the collection of settlement prices from a sequence of LAD problems does not necessarily support the ISO's dispatch (see Section 6.2). In this situation, as discussed in Chapter 3, a generator might have an incentive to deviate from the ISO's dispatch. ISOs typically provide out-of-market payments (usually referred to as "uplift") when necessary to incentivize these generators to follow dispatch instructions, but these payments

[2]The upcoming interval is the the most immediate future interval in the time horizon. It is also called the "binding" interval in some literature. However, to avoid collision with the concept of a binding constraint, we use the word "upcoming".

[3]A set of prices is said to support the ISO's solution if each of the price-taking market participants' profit maximizing decisions align with the ISO's solution.

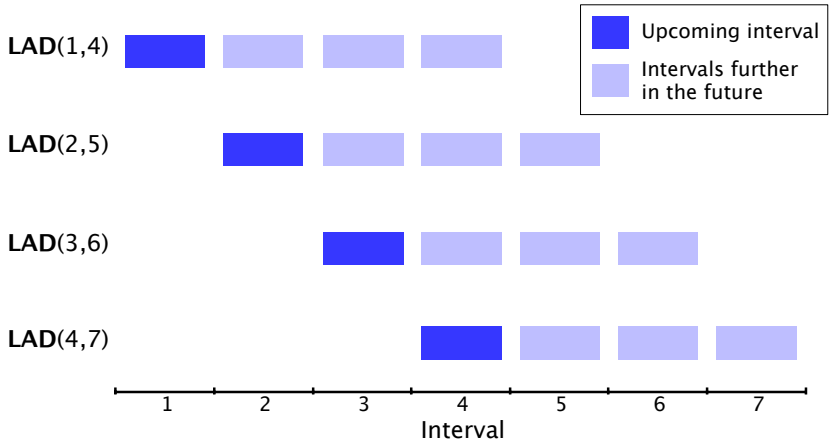


Figure 6.1: A series of LAD problems solved in a rolling implementation of MIRTM. Each look-ahead dispatch problem considers four intervals. For the first problem ($\mathbf{LAD}(1,4)$) that considers a fixed horizon of $t \in [1,4]$, only the dispatch decisions and prices for the upcoming interval (in dark blue) are used for settlement; dispatch decisions and prices for the other three intervals (in light blue) are only advisory.

are non-transparent and discriminatory [34]. In addition to uplift, penalties may apply for generators who deviate from the ISO's instructions. See [77] for a discussion.

Several authors discuss the incentives under MIRTM. For example, Peng and Chatterjee [65] recognize the incentive to deviate in MIRTM but only give a high-level design of their solution. Ela and O'Malley [20] propose cross interval marginal prices (CIMP) where the energy price for the first interval includes the marginal cost of serving load in the subsequent intervals. However, no theoretical or empirical analysis is provided for incentive properties of CIMPs. Moreover, CIMPs may differ from traditional locational marginal prices (LMPs) even when LMPs support the ISO's dispatch solu-

tion. Schiro [74] proposes a multi-settlement market where the ISO settles on all prices in each of the successive LAD problems. This multiple-settlement approach mitigates the incentive problem but complicates the settlement process.

Other lines of relevant research consider a slightly different setting. For example, Thattle et al. [80] study LMPs in LAD-based markets and compare them with prices from a single-interval market. Choi and Xie [13] study the sensitivity of LAD-based markets with respect to data perturbation. However, the two aforementioned studies do not consider incentive issues. Kaye and Outhred [39] address incentive issues in a multi-interval market but do not consider a rolling implementation.

To mitigate the incentive to deviate in MIRTM, a promising pricing method suggested in [35] dualizes system-wide constraints in the past intervals. We show that this existing method gives satisfactory incentives for a fixed finite horizon with perfect foresight.[4] However, in the presence of a rolling horizon, it may produce prices that incentivize a physically infeasible dispatch. Therefore, we propose a new method that instead dualizes binding inter-temporal constraints that link the past and the future. We show that this new method produces appropriate incentives for a fixed finite horizon with perfect foresight. In more practical scenarios where we have a rolling horizon, computational studies suggest that the new method better mitigates incentives

[4]Perfect knowledge about future demand and future offers from market participants.

to deviate than the existing method.

Section 6.2 formulates the LAD problem and illustrates potential incentives to deviate with a small example. Section 6.3 describes the existing pricing method and Section 6.4 presents a new pricing method. Section 6.5 analyzes the theoretical properties of the two pricing methods for a fixed finite horizon. Section 6.6 discusses more practical scenarios where we have a rolling horizon. Section 6.7 describes numerical tests for a system based on the ISO New England grid.

6.2 Multi-Interval Real-Time Market

In this section, we formulate the LAD problem. Using a small example, we show that settlement prices derived from LAD might lead to incentives to deviate for market participants. Throughout this chapter, we treat that all generators as price takers with truthful offers.

6.2.1 Look-Ahead Dispatch

LAD minimizes total incremental cost[5] over the relevant time horizon subject to *system-wide* and *private* constraints. System-wide constraints are coupling constraints that enforce system-wide requirements (e.g., power-balance and transmission constraints). Private constraints include generator output limits and ramp rate restrictions.

[5]The costs do not include fixed (start-up, no-load) costs of resources that are already committed but do include all of the incremental energy costs going forward.

We first define the symbols we use in this chapter:

Sets and indices:

$k \in \mathcal{K}$ Generators.

t Index for time interval.

Parameters:

T_0 Starting interval of the look-ahead horizon.

T End interval of the look-ahead horizon.

d_t System-wide load at time t.

$\overline{p}_k$ Generator k's maximum output level.

$\underline{p}_k$ Generator k's minimum output level.

A_k Coefficient matrix for inter-temporal constraints of generator k.

$\boldsymbol{b}_k$ Constant term for inter-temporal constraints of generator k.

Variables:

p_{kt} Generator k's power output at time t.

λ_t Dual variable associated with the power-balance constraint at interval t.

$\boldsymbol{\mu}_k$ Dual vector associated with the inter-temporal constraints of generator k.

$\boldsymbol{\phi}_k$ Dual vector associated with the output limits of generator k.

Functions:

f_{kt} Generator k's convex cost function at interval t.

We denote the look-ahead dispatch problem as $\mathbf{LAD}(T_0, T)$, where the ordered pair (T_0, T) indicates the starting and ending intervals of the *look-ahead horizon* and $\mathcal{K}$ is the set of generators:[6]

$$\begin{aligned}
\min \quad & \sum_{k\in\mathcal{K}} \sum_{t\in[T_0,T]} f_{kt}(p_{kt}) && && (6.1)\\
\text{over} \quad & p_{kt}, k \in \mathcal{K}, t \in [T_0, T] && && (6.2)\\
\text{s.t.} \quad & -\sum_{k\in\mathcal{K}} p_{kt} + d_t = 0, && \forall t\text{: } \lambda_t && (6.3)\\
& A_k \boldsymbol{p}_k \le \boldsymbol{b}_k, && \forall k\text{: } \boldsymbol{\mu}_k && (6.4)\\
& \underline{\boldsymbol{p}}_k \le \boldsymbol{p}_k \le \overline{\boldsymbol{p}}_k, && \forall k\text{: } \boldsymbol{\phi}_k = (\boldsymbol{\phi}_k^+, \boldsymbol{\phi}_k^-), && (6.5)
\end{aligned}$$

where $\boldsymbol{p}_k = (p_{kT_0}, \ldots, p_{kT})^\intercal \in \mathbb{R}^T$.

We assume that the marginal cost functions f_{kt} for each generator are convex. We minimize total cost. Power-balance constraints (6.3) require total generation to equal total demand; these constraints are the only type of system-wide constraint we consider in the analysis and examples. We discuss extension to the transmission-constrained case in Section 6.4. Constraints (6.5) limit power injection from each generator.

[6]To facilitate our analysis, we assume a starting point T_0 when any loss in the past has been compensated, and when the dispatch trajectory at previous intervals can be ignored.

Since the pricing methods we are about to describe work with any linear inter-temporal constraints, we present a general formulation of these constraints. All private inter-temporal constraints are aggregated into (6.4), which could include ramping constraints, maximum energy constraints, and state-of-charge constraints for energy storage resources. Despite the generality of our formulation, we note that ramping constraints are the only type of inter-temporal constraints we consider in all the numerical examples in this chapter. The formulation of the ramping constraints are shown as follows:

$$p_{kt} - p_{k,t-1} \leq r_k, \, \forall t \tag{6.6}$$

$$p_{k,t-1} - p_{kt} \leq r_k, \, \forall t, \tag{6.7}$$

where r_k is the limit for both the upward and downward ramping of generator k.

The dual variables, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$, and $\boldsymbol{\phi}$, are displayed to the right of their corresponding constraints. If an ISO adopts **LAD** as the pricing problem, λ_t would be the energy price for interval t, with the price for $t = T_0$ being the settlement price and the prices for $t = T_0 + 1, \ldots, T$ being advisory prices. We refer to prices obtained from **LAD** as locational marginal prices (LMPs).[7]

6.2.2 Example 6.1

In this subsection, a simple example shows that the collection of settlement prices from a sequence of **LAD** problems does not necessarily support

[7] In this simplified framework, the LMPs do not vary by location because no transmission constraints are modeled.

Table 6.1: Unit Parameters

Unit	Energy \$/MWh	$\underline{p}_g$ MW	$\overline{p}_g$ MW	Ramping MW/min
1	28	0	100	3
2	30	0	100	4

Table 6.2: Optimal Solution to $\mathbf{LAD}(1, 2)$

t	d_t MW	$p_{1,t}$ MW	$p_{2,t}$ MW	λ_t \$/MWh
1	130	95	35	28
2	155	100	55	**32**

Table 6.3: Optimal Solution to $\mathbf{LAD}(2, 3)$

t	d_t MW	$p_{1,t}$ MW	$p_{2,t}$ MW	λ_t \$/MWh
2	155	100	55	**30**
3	160	100	60	30

the ISO's dispatch.

Example 6.1. Consider a stylized example in which we have two generating units (referred to as units hereafter): Unit 1 is cheaper than Unit 2 but has less ramping capability (see Table 6.1). We consider three 5-minute intervals and a demand vector $\boldsymbol{d} = (130, 155, 160)$. We assume that each LAD problem models two future intervals. The initial power outputs ($t = 0$) of the two units are $p_{10} = 95$ MW and $p_{20} = 35$ MW. Table 6.2 and Table 6.3 show the solutions of the sequential dispatch problems $\mathbf{LAD}(1, 2)$ and $\mathbf{LAD}(2, 3)$.

The ramping constraint between $t = 1$ and $t = 2$ for Unit 2 is binding in $\mathbf{LAD}(1, 2)$, reducing the price in $t = 1$ and raising the price in $t = 2$. Following currently implemented ISO approaches, only the price for $t = 1$ from $\mathbf{LAD}(1, 2)$ is used for settlement. Unit 2 loses \$5.83 at $t = 1$, but believes that it will recover its losses due to the high advisory price for $t = 2$. However, the subsequent dispatch problem $\mathbf{LAD}(2, 3)$ treats Unit 2's $t = 1$ loss as sunk and identifies a settlement price of \$30/MWh for $t = 2$.[8] Unit 2's loss is not compensated by energy prices.

6.2.3 Incentive to Deviate

Given an instance of $\mathbf{LAD}(T_0, T)$, suppose the ISO obtains an optimal dispatch vector $\boldsymbol{p}^*$ and energy price vector $\boldsymbol{\lambda}^*$. Since **LAD** is a convex problem, the prices across multiple intervals (vector $\boldsymbol{\lambda}^*$) as a whole support the ISO's welfare-maximizing solution $\boldsymbol{p}^*$. That is, assuming that generator k is a price-taker, the ISO-determined decision $\boldsymbol{p}_k^*$ is individually rational for generator k. Mathematically, this means that $\boldsymbol{p}_k^*$ is an optimal solution to generator k's profit maximization problem:

$$
\begin{aligned}
w_k(\boldsymbol{\lambda}^*) = \quad & \max_{\boldsymbol{p}_k} \quad (\boldsymbol{\lambda}^*)^\intercal \boldsymbol{p}_k - \sum_{t \in [T_0, T]} f_{kt}(p_{kt}) && (6.8) \\
& \text{s.t.} \quad A_k \boldsymbol{p}_k \leq \boldsymbol{b}_k, && (6.9) \\
& \qquad \underline{\boldsymbol{p}}_k \leq \boldsymbol{p}_k \leq \overline{\boldsymbol{p}}_k, && (6.10)
\end{aligned}
$$

[8] Note that dual degeneracy exists for $\mathbf{LAD}(2, 3)$ and λ_2^* can be any value in $[30, \infty]$. We believe that a commercial solution algorithm is likely to find the price \$30/MWh.

where w_k is the value function of this problem denoting the maximum profit.

As long as the ISO commits to settlement for all intervals $T_0, \ldots, T$ at the corresponding prices λ_t^*, losses incurred due to binding inter-temporal constraints will be compensated by high prices in other intervals. For instance, in Example 6.1, if the ISO were to commit to the price vector $(28, 32)$ computed from $\mathbf{LAD}(1, 2)$, there would be no incentive to deviate.

However, as mentioned in the introduction, CAISO and NYISO's MIRTM implementations only settle at the upcoming interval's price in each of the successive LAD problems. Subsequent LAD problems treat losses incurred in previous intervals as sunk, leading to incentives to deviate. This can be seen in Example 6.1: the settlement price for $t = 2$ is lower than the advisory price for $t = 2$ published in the previous interval. Had Unit 2 known the $t = 2$ settlement price, it would have generated 0 MW at $t = 1$ instead of the ISO's dispatch. In practice, the ISO pays each generator who has an incentive to deviate an out-of-market payment as will be discussed in Section 6.6.2.

6.3 An Existing Pricing Method

Intuitively, we can create a separate pricing problem that incorporates losses from past intervals. In this approach, the dispatch problem (**LAD**) determines the dispatch quantities and a separate *pricing problem* identifies the prices.[9] We note that such a separation of dispatch and pricing problems

[9]The dispatch solutions in the pricing problem are discarded.

has been a common practice in US wholesale electricity markets. For example, ISO New England, Midcontinent ISO, and NYISO have separate dispatch and pricing runs in their real-time markets.

In this section, we describe an existing pricing proposal that employs such a separate pricing problem. Using a small example, we show that this method might produce prices that incentivize a physically infeasible solution.

6.3.1 Price-Preserving Multi-Interval Pricing

To provide better incentives in MIRTM, Hogan [35] suggests that the pricing problem constrain prices in the past intervals to be equal to the realized prices. Although a detailed implementation of this idea is not provided in [35], the desired result can be achieved by making the following changes to **LAD**: 1) keep past dispatch decisions as variables, and 2) dualize system-wide constraints corresponding to the past intervals with a penalty equal to realized prices.[10] Intuitively, historical loss is incorporated through the penalties that equal the realized prices.

We denote this pricing problem by $\mathbf{PMP}(T_0, \hat{t}, T)$, where the ordered triplet $(T_0, \hat{t}, T)$ comprises three numbers indicating the earliest past interval considered, the upcoming interval, and the ending interval of the horizon, respectively. That is, we "look back" $\hat{t} - T_0$ intervals. Let λ_t^* be the realized settlement prices for $t \in \{T_0, \ldots, \hat{t} - 1\}$.

[10] Additional details of this idea were discussed in personal communications between ISO New England and William Hogan.

PMP$(T_0, \hat{t}, T)$ (Price-preserving Multi-interval Pricing) is formulated as follows:

$$\begin{aligned}
\min \quad & \sum_{k\in\mathcal{K}} \sum_{t\in[T_0,T]} f_{kt}(p_{kt}) + \sum_{t=T_0}^{\hat{t}-1} \lambda_t^*(-\sum_k p_{kt} + d_t) && (6.11)\\
\text{over} \quad & p_{kt}, k \in \mathcal{K}, t \in \{T_0, \ldots, T\} && (6.12)\\
\text{s.t.} \quad & -\sum_k p_{kt} + d_t = 0, \forall t \in \{\hat{t}, \ldots, T\} \qquad : \lambda_t && (6.13)\\
& A_k \boldsymbol{p}_k \le \boldsymbol{b}_k, \forall k \qquad : \boldsymbol{\mu}_k && (6.14)\\
& \underline{\boldsymbol{p}}_k \le \boldsymbol{p}_k \le \overline{\boldsymbol{p}}_k, \forall k \qquad : \boldsymbol{\phi}_k && (6.15)
\end{aligned}$$

where $\boldsymbol{p}_k = (p_{k1}, \ldots, p_{kT})^\intercal \in \mathbb{R}^T$. We use the same symbols $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\phi})$ as in **LAD** for the dual variables in this optimization problem. Note that the dispatch decisions for the past intervals are kept as variables, and that all private constraints are preserved for those intervals.

We set the length of the time horizon $T - T_0$ of the pricing problem to be equal to that of the LAD problem. The length of the horizon impacts the operational cost and price volatility of the MIRTM. The choice of $T - T_0$ is beyond the scope of this chapter. We refer the interested readers to [80] for a discussion.

Applying PMP to Example 6.1, we solve **PMP**$(1, 2, 3)$ at $t = 2$ with a $t = 1$ penalty of \$28/MWh. The resulting $t = 2$ price from PMP is \$32/MWh. The final settlement price vector $(\pi_1, \pi_2) = (28, 32)$ gives neither unit an incentive to deviate.

Table 6.4: Unit Parameters

Unit	Energy \$/MWh	$\underline{p}_g$ MW	$\overline{p}_g$ MW	Ramping MW/min
1	28	0	100	3
2	30	0	100	4
3	40	0	100	5

6.3.2 Example 6.2

Although PMP works well for Example 6.1, we show another example for which PMP identifies prices that incentivize an infeasible dispatch.

Example 6.2. Consider the addition of an expensive Unit 3 to Example 6.1 (see Table 6.4). We also consider an additional interval $t = 4$, and augment the demand vector to $\boldsymbol{d} = (130, 155, 180, 180)$. Initially, the power output of each generator is $(p_{10}, p_{20}, p_{30}) = (95, 35, 0)$. We assume that the MIRTM considers two future intervals. Table 6.5 shows:

- the realized dispatch obtained by sequential solutions to $\mathbf{LAD}(1, 2)$, $\mathbf{LAD}(2, 3)$, and $\mathbf{LAD}(3, 4)$;
- the settlement prices obtained by sequential **LAD** problems (denoted by LMP from here on) and those obtained by **PMP**.

Surprisingly, the $t = 3$ settlement price produced by PMP is \$34/MWh, lower than the marginal cost of Unit 3. As a result, Unit 3 would be better off shutting down and has an incentive to deviate.

To investigate this pricing, Table 6.6 shows the solution to $\mathbf{PMP}(1, 3, 4)$.

Table 6.5: Realized Dispatch and Settlement Prices from Rolling Implementations

t	d_t	$p_{1,t}$	$p_{2,t}$	$p_{3,t}$	LMP	PMP
	MW	MW	MW	MW	\$/MWh	\$/MWh
1	130	95	35	0	28	28
2	155	100	55	0	28	28
3	180	100	75	5	40	**34**

Table 6.6: Optimal Solution to $\mathbf{PMP}(1, 3, 4)$

t	d_t	$p_{1,t}$	$p_{2,t}$	$p_{3,t}$	λ_t
	MW	MW	MW	MW	\$/MWh
1	130	85	40	0	-
2	155	100	**60**	0	-
3	180	100	**80**	**0**	34
4	180	100	80	0	30

In this table, the $t = 3$ dispatch for Unit 3 is 0 MW, indicating that (from a pricing perspective) Unit 3 shouldn't be online or set price. Furthermore, because PMP keeps past dispatch decisions as variables, the counter-factual dispatch trajectory in PMP can differ significantly from the realized (LAD-instructed) solution (see Fig. 6.2). As a result, the dispatch decisions for Unit 2 from $\mathbf{PMP}(1, 3, 4)$ are physically infeasible: at $t = 3$, the 80 MW dispatch point from $\mathbf{PMP}(1, 3, 4)$ violates Unit 2's ramping limit from its realized 55 MW dispatch at the previous interval.

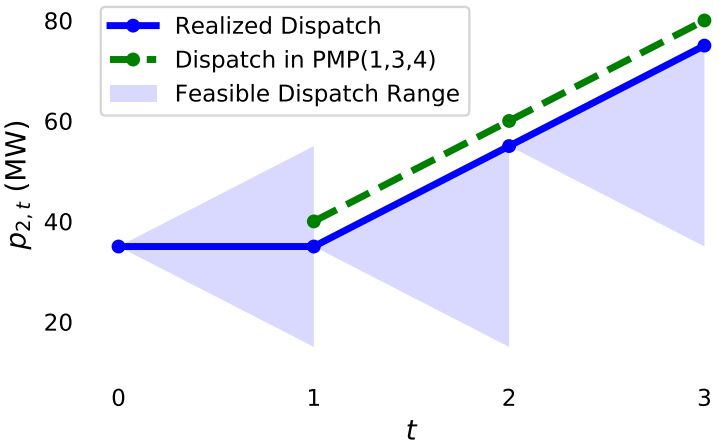


Figure 6.2: Assume that Unit 2 follows the LAD instructions (described in Table 6.5 and shown as the solid line here). The feasible dispatch range at t is determined by technical constraints relative to the realized (LAD-instructed) dispatch at $t-1$. Suppose the ISO adopts **PMP** as the pricing problem. At $t = 3$, **PMP**$(1, 3, 4)$ is solved; its dispatch decisions for Unit 2 are shown as the dotted line. The $t = 3$ dispatch of 80 MW from PMP violates Unit 2's ramp rate limit based on the realized dispatch $p_{22} = 55$. **PMP**$(1, 3, 4)$ incentivizes this infeasible dispatch.

6.4 Proposed Pricing Method

Since **PMP** keeps past dispatch decisions as variables, its dispatch decisions can be infeasible relative to realized unit output (e.g., violating ramping constraints, or dispatching a unit off that is needed to meet demand). As a result, prices produced by PMP might lead to large uplift payments.

To overcome this drawback of PMP, this section proposes a different pricing method that fixes past dispatch decisions to their realized values. Instead of dualizing system-wide constraints, the new method relaxes intertemporal constraints that link the past to the future. Price consistency is achieved by setting the penalty for these relaxed constraints to their corre-

sponding optimal dual variables from the past.

We let $\boldsymbol{\mu}_k^*$ denote the realized optimal dual vector associated with generator k's inter-temporal constraints in **LAD**. Let $A_k = [A_k'; A_k''; A_k''']$, $\boldsymbol{b}_k = [\boldsymbol{b}_k'; \boldsymbol{b}_k''; \boldsymbol{b}_k''']$, and $\boldsymbol{\mu}_k = [\boldsymbol{\mu}_k'; \boldsymbol{\mu}_k''; \boldsymbol{\mu}_k''']$[11] be a partition of the inter-temporal constraints (as well as their corresponding coefficients and dual variables) such that:

1. constraints $A_k'\boldsymbol{p}_k \leq \boldsymbol{b}_k'$ only involve the dispatch decisions in the past with respect to $\hat{t}$ (i.e., each row of A_k' has all zero entries in the last $T - \hat{t} + 1$ columns);
2. constraints $A_k''\boldsymbol{p}_k \leq \boldsymbol{b}_k''$ involve both past and future decisions with respect to $\hat{t}$ (i.e. each row of A_k'' has at least one nonzero entry in the first $\hat{t} - 1$ columns, and at least one nonzero entry in the last $T - \hat{t} + 1$ columns);
3. constraints $A_k'''\boldsymbol{p}_k \leq \boldsymbol{b}_k'''$ only involve dispatch decisions in the future with respect to $\hat{t}$ (i.e., each row of A_k''' has all zero entries in the first $\hat{t} - 1$ columns).

Let $\boldsymbol{p}_k'^* = (p_{k1}, \ldots, p_{k,\hat{t}-1})$ be the realized dispatch decisions for generator k in the past. Let $\boldsymbol{p}_k''' = (p_{k\hat{t}}, \ldots, p_{kT})^\intercal \in \mathbb{R}^{T-\hat{t}+1}$ be the dispatch variables for the future. Note that $[\boldsymbol{p}_k'^*; \boldsymbol{p}_k''']$ represents the complete decision vector, with past decisions fixed at their realized values and future decisions considered variables. We denote the following optimization problem by $\mathbf{CMP}(\hat{t}, T)$

[11] The semicolon indicates that the matrices/vectors are stacked vertically.

(Constraint-preserving Multi-interval Pricing), where the ordered pair indicates the upcoming interval and the ending interval of the look-ahead horizon:

$$\begin{aligned}
\min \quad & \sum_{k\in\mathcal{K}} \sum_{t\in[\hat{t},T]} f_{kt}(p_{kt}) + (\boldsymbol{\mu}_k''^*)^\intercal A_k''[\mathbf{0}; \boldsymbol{p}_k'''] && && (6.16)\\
\text{over} \quad & p_{kt}, k \in \mathcal{K}, t \in \{\hat{t}, \ldots, T\} && && (6.17)\\
\text{s.t.} \quad & -\sum_k p_{kt} + d_t = 0, && \forall t\text{: } \lambda_t && (6.18)\\
& A_k''[\boldsymbol{p}_k'^*; \boldsymbol{p}_k'''] \le \boldsymbol{b}_k'', && \forall k\text{: } \boldsymbol{\mu}_k'' && (6.19)\\
& A_k'''[\mathbf{0}; \boldsymbol{p}_k'''] \le \boldsymbol{b}_k''', && \forall k\text{: } \boldsymbol{\mu}_k''' && (6.20)\\
& \underline{\boldsymbol{p}}_k \le \boldsymbol{p}_k''' \le \overline{\boldsymbol{p}}_k, && \forall k\text{: } \boldsymbol{\phi}_k && (6.21)
\end{aligned}$$

Reusing notation from **LAD** again, we use the same symbols $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\phi})$ for the dual variables. Note that we have replaced the first $\hat{t}-1$ entries of the vector $[\boldsymbol{p}_k'^*; \boldsymbol{p}_k''']$ by zero in (6.16) and (6.20). This is to indicate that only future decisions $\boldsymbol{p}_k'''$ matter in these equations.

Since **CMP** fixes past dispatch decisions to their realized values, and since CMP dualizes inter-temporal constraints that link the past to the future, we no longer have to represent past intervals. This is an important difference between **CMP** and **PMP**.

The second term in the objective function (6.16) results from the dualization of inter-temporal constraints that link the past to the future. There remain two other sets of inter-temporal constraints: (6.19) represents coupling between past fixed dispatches and future decisions; (6.20) couple only future decisions.

Finally, we note that **CMP** can be easily extended to consider linearized transmission constraints. Since **CMP** does not modify the system-wide constraints, we can apply **CMP** to the transmission-constrained case without any special treatment. The locational prices can be derived in the standard way (e.g., see [46]).

For **PMP**, the original formulation in [35] dualizes only power balance constraints in the past. To constrain all locational prices in the past to be equal to the realized values, we can dualize each system-wide constraints (power balance constraint and transmission constraint) in the past, with a penalty equal to its realized optimal dual variable.

6.5 Fixed Finite Horizon

In this section, we first propose a fixed-horizon framework that facilitates the analysis of pricing methods. We show that both **PMP** and **CMP** produce satisfactory prices in this ideal framework. However, the two pricing methods lead to different incentives in the scenario where there is a moving time horizon (depicted in Fig. 6.1). This more practical scenario will be considered in the next section.

6.5.1 Fixed Finite Horizon and Consistent Prices

As a starting point of our analysis, this section considers a fixed finite horizon $t = \{1, \ldots, T\}$ as illustrated in Fig. 6.3. For ease of analysis, we treat $t = 1$ as the beginning of time, where we assume that any loss in the past

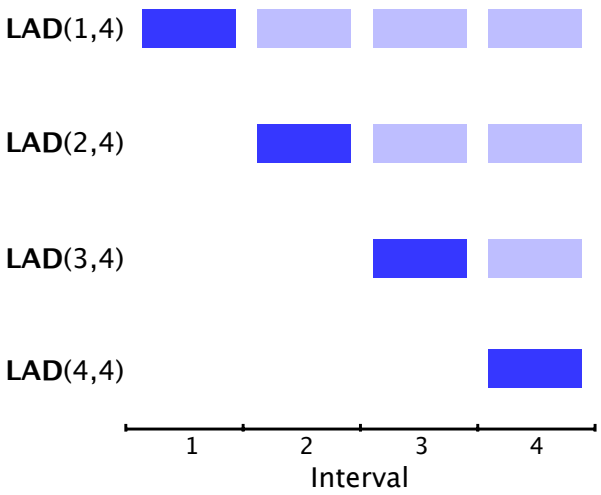


Figure 6.3: A series of LAD problems solved for a fixed finite horizon of four intervals. As a starting point of our analysis, we ignore the new intervals that would have been included in the more practical situation where there is a moving horizon.

for every generator has been compensated. Therefore, **LAD** is the pricing problem for $t = 1$. Starting with $t = 2$, there might be losses from earlier intervals. A separate pricing problem identifies prices for $t \geq 2$.

With the assumption of a fixed finite horizon, no new information (e.g., demand and offers from market participants at a future interval) is introduced as time progresses. Given no change in demand and supply in this setting, it is desirable for the settlement price and all advisory prices for a given interval to be identical. To formalize this idea, we make the following assumption.

Assumption 6.1 (Perfect Foresight)**.** The demand forecast is perfect and the market participants' offered parameters do not change as time progresses.

We then have the following definition:

Definition 6.1 (Consistent Prices)**.** Given Assumption 6.1, a pricing method for MIRTM gives consistent prices if successive pricing problems produce the same price for the same interval.

For Example 6.1, if a pricing method produces consistent prices, then the advisory price for $t = 2$ obtained at $t = 1$ would equal the settlement price for $t = 2$. If we use **LAD** as the pricing problem, then we lose price consistency because the advisory price for $t = 2$ ($32/MWh) is different from the settlement price ($30/MWh) for the same interval.

6.5.2 Analysis of PMP and CMP

We next show that both **PMP** and **CMP** produce consistent prices for a fixed finite horizon with perfect foresight.

Theorem 6.1. *Let $\boldsymbol{\lambda}^*$ be an optimal dual vector associated with constraints* (6.3) *in* ***LAD***$(1, T)$*. For every $\hat{t} \in \{2, \ldots, T\}$, $\left(\left(\lambda^*_{\hat{t}}, \ldots, \lambda^*_T\right), \boldsymbol{\mu}^*, \boldsymbol{\phi}^*\right)$ is an optimal dual solution to* ***PMP***$(1, \hat{t}, T)$ *given Assumption 6.1.*

PMP dualizes part of the constraints in **LAD** with a penalty equal to the optimal dual variables. Theorem 6.1 is implied by the strong duality of **LAD**.

Proof. Let $(\boldsymbol{p}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, \boldsymbol{\phi}^*)$ denote an optimal primal-dual solution to $\mathbf{LAD}(1, T)$, where $\boldsymbol{p}^* \in \mathbb{R}^{|K| \times T}$ is a vector that collects all primal solutions. Since $\mathbf{LAD}(1, T)$ is a convex optimization problem, and since $\mathbf{PMP}(1, \hat{t}, T)$ is only different

from $\mathbf{LAD}(1,T)$ in that some constraints are dualized with a penalty equal to corresponding optimal dual variables, strong duality holds between these two optimization problems. In addition, $\boldsymbol{p}^*$ is also a minimizer of $\mathbf{PMP}(1,\hat{t},T)$ [9, Section 5.5].

Convexity implies that KKT conditions are necessary and sufficient for both optimization problems. With identical primal optimal solutions, comparing the KKT conditions of the two optimization problems leads to the conclusion that

$$(\boldsymbol{p}^*, (\lambda_{\hat{t}}^*, \ldots, \lambda_T^*), \boldsymbol{\mu}^*, \boldsymbol{\phi}^*)$$

is also an optimal primal-dual solution to $\mathbf{PMP}(1,\hat{t},T)$. □

Recall that the solution of $\mathbf{LAD}(1,T)$ gives a binding settlement price for $t = 1$ and advisory prices for $t \in [2,T]$. Theorem 6.1 implies that these advisory prices can be reproduced by the sequential solutions of **PMP** given perfect foresight. Thus, price consistency holds.

We have the following theorem implying that **CMP** produces consistent prices for a fixed finite horizon.

Theorem 6.2. *Let $\boldsymbol{\lambda}^*$ be an optimal dual vector associated with the power-balance constraints in $\boldsymbol{LAD}(1,T)$. Given $\hat{t} \in \{2,\ldots,T\}$, under Assumption 6.1, there exists an optimal dual solution $(\boldsymbol{\lambda}^{**}, \boldsymbol{\mu}^{**}, \boldsymbol{\phi}^{**})$ to $\boldsymbol{CMP}(\hat{t},T)$ such that $\lambda_t^{**} = \lambda_t^*$ for all $t \in \{\hat{t},\ldots,T\}$.*

Proof. First, consider an undecoupled version of **CMP** where we only dualize

inter-temporal constraints that link the past and the future, and where we have not yet removed the decision variables for the past intervals:

$$\begin{aligned}
\min \quad & \sum_{k\in\mathcal{K}} f_k(\boldsymbol{p}_k) + (\boldsymbol{\mu}_k^{\prime\prime,*})^\intercal (A_k''\boldsymbol{p}_k - b_k'') && && (6.22)\\
\text{over} \quad & p_{kt}, k \in \mathcal{K}, t \in \{1,\ldots,T\} && && (6.23)\\
\text{s.t.} \quad & -\sum_k p_{kt} + d_t = 0, && \forall t\text{: } \lambda_t && (6.24)\\
& A_k'\boldsymbol{p}_k \le \boldsymbol{b}_k', && \forall k\text{: } \boldsymbol{\mu}_k' && (6.25)\\
& A_k'''\boldsymbol{p}_k \le \boldsymbol{b}_k''', && \forall k\text{: } \boldsymbol{\mu}_k''' && (6.26)\\
& \underline{\boldsymbol{p}}_k \le \boldsymbol{p}_k \le \overline{\boldsymbol{p}}_k, && \forall k\text{: } \boldsymbol{\phi}_k, && (6.27)
\end{aligned}$$

Similar to the proof of Theorem 6.1, because of strong duality, an optimal primal-dual solution to $\mathbf{LAD}(1,T)$ is also optimal for this undecoupled problem.

A sub-vector of the optimal primal-dual solution to the undecoupled problem above is also optimal for $\mathbf{CMP}(\hat{t},T)$, because the two problems only differ in the following ways that do not alter the optimal solution:

1. separability resulting from dualized inter-temporal constraints allows us to discard decision variables for the past intervals, as well as their corresponding items in the objective function;
2. constant term $(\boldsymbol{\mu}_k^{\prime\prime,*})^\intercal b_k''$ is discarded in $\mathbf{CMP}(\hat{t},T)$;
3. constraints (6.25) only involve decision variables for the past intervals and are discarded in $\mathbf{CMP}(\hat{t},T)$;
4. constraints (6.19) are added to $\mathbf{CMP}(\hat{t},T)$; these constraints are redundant because any optimal solution to the undecoupled problem

satisfies these constraints.

□

6.6 Rolling Horizon

Section 6.5 has shown that both PMP and CMP produce consistent prices for a fixed finite time horizon with perfect foresight. Prices produced by either of these two methods incentivize the ISO's welfare-maximizing decisions throughout the fixed time horizon.

In practice, the look-ahead horizon moves forward with time. Accordingly, we solve the proposed pricing problems on a rolling basis just as the dispatch problems. In this general case, neither **PMP** nor **CMP** guarantees price consistency since new variables and constraints corresponding to the new future interval alter the optimality conditions.

By revisiting Example 6.2 we show that, unlike **PMP**, the dispatch decisions in **CMP** are feasible relative to realized unit output. We then describe two types of out-of-market payments as quantitative measures of the incentives provided by energy prices. We will use these measures in our computational studies to examine how well the proposed pricing methods mitigate the incentive to deviate.

Table 6.7: Optimal Solution to **CMP**$(3, 4)$

t	d_t MW	$p_{1,t}$ MW	$p_{2,t}$ MW	$p_{3,t}$ MW	λ_t \$/MWh
3	180	100	**75**	**5**	40
4	180	100	80	0	30

6.6.1 Revisiting Example 6.2

Recall that Example 6.2 has four intervals, with a rolling look-ahead horizon of length two. As the horizon moves, the ramping constraint for Unit 2 remains binding and PMP fails to achieve price consistency. Moreover, because PMP keeps past dispatch decisions as variables, the obtained prices incentivize an infeasible solution and Unit 3 has an incentive to deviate.

Table 6.7 show the optimal solutions to **CMP** solved at $t = 3$. We can see that the dispatch variable for Unit 2 at $t = 3$ is 75 MW, a value that is feasible due to constraints (6.19). The set of settlement prices under CMP is $(28, 28, 40)$, providing incentives for all generators to follow the ISO's dispatch.

In the practical scenario with a rolling horizon, the dispatch decisions in **CMP** can often be more reasonable than those in **PMP** because past dispatch decisions are fixed to their realized values. These realized dispatch decisions interact with future dispatch decisions through constraints (6.19), restricting the future dispatch variables to values that are technically feasible given the earlier dispatch decisions.[12] We also note that **CMP** is smaller in size than

[12]They are redundant when we have perfect foresight of a fixed horizon. In the general case

PMP since it does not represent past intervals.

6.6.2 Uplift Payment as a Measure of Incentive

Although price consistency cannot be guaranteed in the general case, it is still desirable to have reasonable prices that incentivize the ISO's optimal dispatch as much as possible. One measure of the effectiveness of price incentives is the amount of out-of-market (i.e., uplift) payments needed to compensate the generators. In general, there are two primary uplift types that an ISO may calculate [75].

Lost Opportunity Cost One type of uplift payment covers a unit's *lost opportunity cost* (LOC), the gap between its maximum possible profit and the actual profit obtained by following the ISO's dispatch instruction. In MIRTM, we adopt an *ex-post* definition of LOC where a unit's maximum possible profit for a specified period of time (e.g., a commitment cycle) is calculated after the publication of all settlement prices within the given time window. Given a vector of settlement prices over relevant intervals $\boldsymbol{\lambda}^*$ and an ISO's optimal dispatch decision $\boldsymbol{p}_k^*$, LOC is:

$$L_k(\boldsymbol{\lambda}^*, \boldsymbol{p}_k^*) = w_k(\boldsymbol{\lambda}^*) - ((\boldsymbol{\lambda}^*)^\intercal \boldsymbol{p}_k^* - f_k(\boldsymbol{p}_k^*)), \tag{6.28}$$

where the first term on the right hand side is the maximum profit computed using the profit maximization problem (6.8)–(6.10), and the second term is the

when we have a rolling horizon with imperfect foresight, these constraints can be binding.

actual profit made by following the ISO's dispatch. Since the ISO's decision $\boldsymbol{p}^*$ is a feasible but not necessarily optimal solution to problem (6.8)–(6.10), LOC is non-negative.

If prices are too high, a generator might have an incentive to generate more than the ISO's dispatch; if prices are too low, there might be an incentive to under-generate. Both scenarios lead to positive LOC. When prices support the ISO's dispatch, LOC is zero.

Make-Whole Payment Another type of uplift payment that most ISOs adopt is *make-whole payment* (MWP). This payment guarantees that no unit that is economically committed by the ISO loses money over a specified period of time (e.g., a commitment cycle). Given a vector of settlement prices $\boldsymbol{\lambda}^*$ and ISO's optimal dispatch vector $\boldsymbol{p}_k^*$, MWP is:

$$M_k(\boldsymbol{\lambda}^*, \boldsymbol{p}_k^*) = \max\{0, f_k(\boldsymbol{p}_k^*) - (\boldsymbol{\lambda}^*)^\intercal \boldsymbol{p}_k^*\}, \tag{6.29}$$

where the second argument of the max operator is the loss (negative profit) generator k incurs by following the ISO's dispatch. When the loss is positive, $M_k > 0$; when the loss is negative (i.e., profit), $M_k = 0$.

If prices are too low, a generator that follows the ISO's dispatch might have a loss, leading to positive MWP. Higher prices generally lead to lower MWP.

Unlike energy prices, uplift payments are unit-specific and thus discriminatory. These private side payments make it harder for a potential entrant

to determine if entry would be profitable, particularly if the uplift payments are not disclosed publicly. Therefore, smaller uplift payments imply a more transparent market where energy prices better incentivize the ISO's dispatch.

We note that while make-whole payments exist in most ISOs, only ISO-NE pays the aforementioned type of LOC among the US ISOs. In the following computational studies, the two types of uplift serve as a quantitative measure of dispatch-following incentives with lower uplift indicating better market incentives.

6.7 Computational Results

We implement three pricing methods (**LAD**, **PMP**, and **CMP**) on a personal computer with a 2.2-GHz quad-core CPU and 16 GB of RAM. We refer to the prices obtained through **LAD** as LMPs, and name the other two types of prices after the pricing model. The optimization problems are solved by GUROBI 6.5 [31].

We consider scenario 1 of [40], which is based on structural attributes and data from ISO New England. We use linear cost functions, no-load costs,[13] and constant marginal cost terms but ignore start-up cost and the quadratic cost terms in the original data. Minimum generation levels for the units are not specified in the original data, so we let $\underline{p} = 0.8\overline{p}$ for each nuclear plant, $\underline{p} = 0.6\overline{p}$ for each coal-fired unit, and $\underline{p} = 0$ for the rest.

[13] Quasi-fixed costs incurred to keep a generator online that is independent of the amount of energy generated.

We assume the following setting for MIRTM:

- an interval length of five minutes;
- a look-ahead length of five intervals,
- five most recent past intervals included in **PMP**;
- ramping constraint is the only type of inter-temporal constraint we consider.

6.7.1 Ramping Up

Typically, ramping constraints are binding when the increase of demand from one interval to the next is high. Therefore, we focus on the hour-long horizon between the sixth and the seventh hour of the original dataset,[14] where there is a 2046 MW demand increase. Since we assume an interval length of five minutes, we use linear interpolation to determine the 11 demand values within the hour. We then add random Gaussian noise to each demand value. The Gaussian noise has zero mean and a standard deviation of 2% of the interpolated demand. We first focus on one sample of the load curve (Fig. 6.4).

There are 47 units committed within this hour. The commitment decisions are made by solving a 96-hour unit commitment problem. The total generation cost from LAD is $\$6.33 \times 10^6$. We next study the performance of the three pricing methods on this sample.

[14]The original data contain four-day hour-by-hour demand information designed for the day-ahead market.

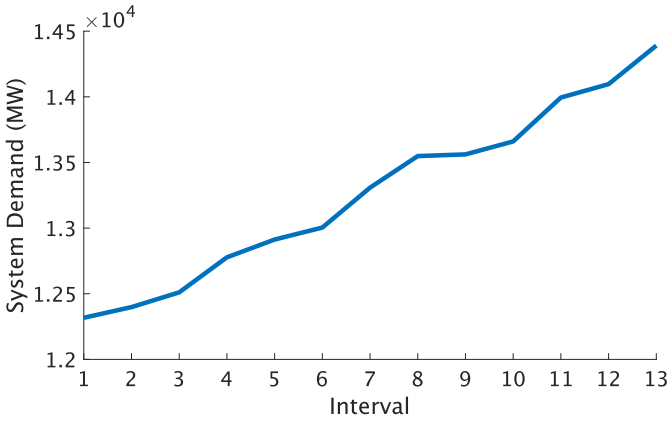


Figure 6.4: A sample of a ramping-up load curve.

Fig. 6.5 shows settlement and advisory prices under the three pricing methods. Since the look-ahead length is five intervals, the last dispatch problem is $\mathbf{LAD}(9, 13)$. As a result, we obtain nine settlement prices. Since we begin with intervals $t = 1$, there are $t - 1$ advisory prices for each interval $t \leq 4$; for all other intervals, there are four advisory prices. We use a shaded area to show the range of prices (settlement and advisory) for a given interval. The total energy payments (excluding uplift payments) from load under LMP, CMP, and PMP are respectively 3.15×10^7, 3.26×10^7, and 2.23×10^7. For this example, there is a welfare transfer from consumers to suppliers if we use CMP instead of LMP.

As seen in Fig. 6.4, a high ramping requirement happens between $t = 6$ and $t = 7$. The corresponding ramping constraints for multiple expensive units are binding, causing the price at $t = 6$ to decrease and the price at $t = 7$ to increase. For instance, in $\mathbf{LAD}(6, 10)$, the advisory price for $t = 6$ is \$52/MWh,

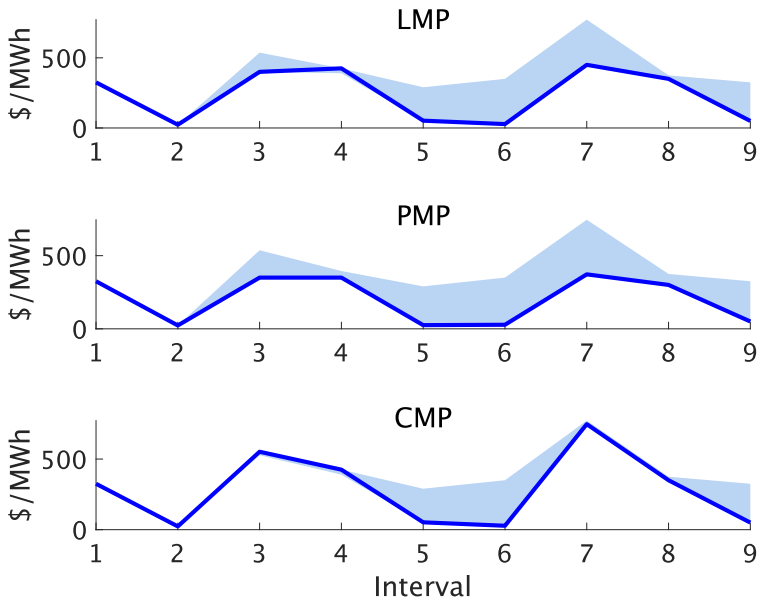


Figure 6.5: Settlement and advisory prices under different pricing methods. For each interval, the solid line shows the settlement prices; the area between maximum and minimum prices (including settlement and advisory) is shaded. Values at fractional points on the interval axis are linearly interpolated. We can see that the high advisory prices for $t = 7$ does not realize under LMP, whereas under CMP the settlement price equals the advisory prices.

lower than the highest marginal price of the dispatched units, while the advisory price for $t = 7$ is \$598/MWh, higher than the highest marginal price of the dispatched units. As time progresses to $t = 7$, $\mathbf{LAD}(7, 11)$ treats historical loss as sunk. Consequently, the settlement LMP for $t = 7$ is \$400/MWh, lower than the advisory prices for that interval. This is similar to, but more extreme than, what we saw in Example 6.1. In contrast, the advisory prices are equal to the settlement price at $t = 7$ when we use **CMP**.

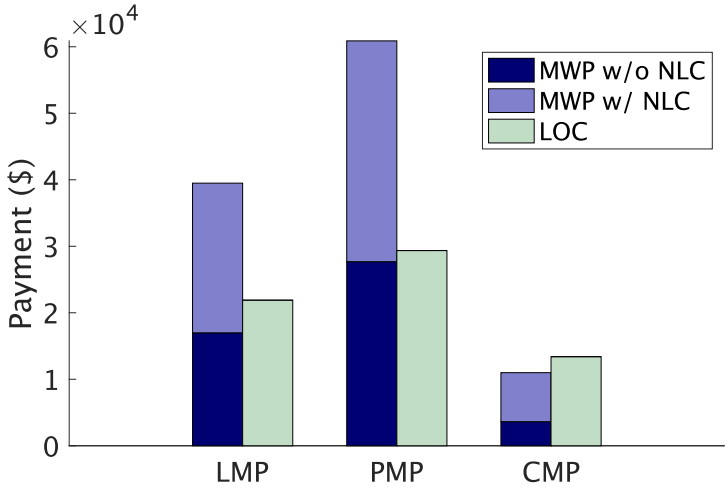


Figure 6.6: Total make-whole payments (with and without considering no-load costs, NLC) and lost opportunity costs under different pricing methods.

We also note that although **PMP** incorporates historical loss, the resulting prices are often lower than LMP. This can be explained by the potentially infeasible dispatch decisions in **PMP**. For example, multiple expensive units are dispatched at $t = 8$ to meet the increase in demand. In **PMP** $(3, 8, 12)$ however, similar to what happened in Example 6.2, the dispatch decisions of some units at $t = 8$ are infeasible based on the realized dispatch, leading to low prices.

To examine the effectiveness of the prices in incentivizing the ISO's decisions, we compute the make-whole payments and lost opportunity costs resulting from different prices, shown in Fig. 6.6. Because no-load cost is lumpy and cannot be incorporated in marginal pricing, we compute two types of make-whole payments: one that includes no-load costs, and one that does

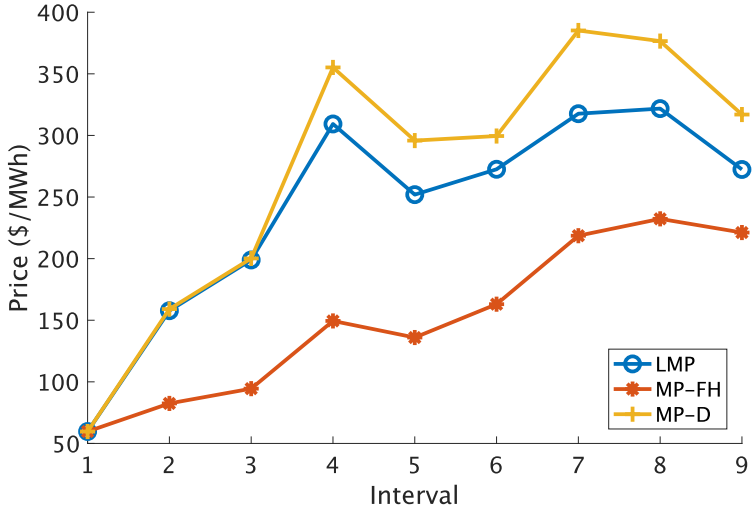


Figure 6.7: Mean settlement prices under different pricing methods.

not. We see that CMP results in the least MWP and the least LOC.

6.7.2 More Samples

We sample 50 load curves and analyze the performance of three pricing methods on these samples. The samples are created by adding the same Gaussian noise as in the previous subsection. Fig. 6.7 shows the settlement prices of each interval averaged over all the samples. We can see that CMPs are on average higher than LMPs for most intervals, whereas PMPs are the lowest. These observations are consistent with what we have seen in Section 6.7.1.

Since CMPs are on average higher than LMPs, we expect CMPs to reduce MWP. However, if prices are too high, a generator might have incentives to over-generate, resulting in high LOC. Fig. 6.8 shows both type of uplift

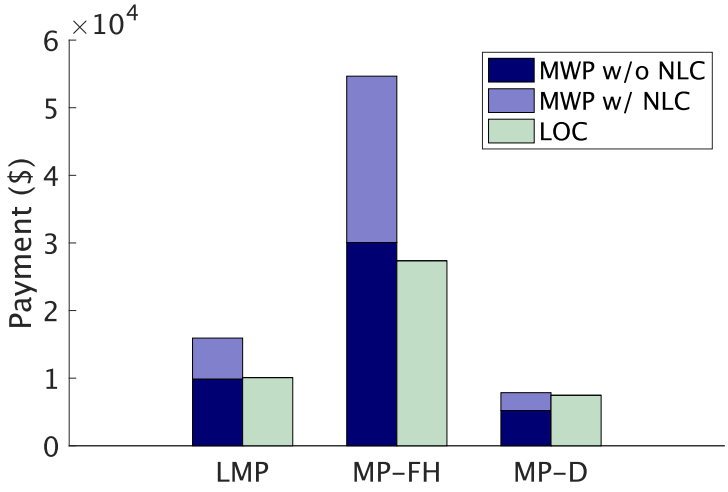


Figure 6.8: Mean total uplift payments under different pricing methods.

payments under the three pricing methods. We can see that CMPs best incentivize the ISO's dispatch, since they lead to the least MWP and LOC on average.

Part V

Conclusions

Chapter 7

Conclusions

We conclude this dissertation by summarizing our findings and describing future research opportunities. We present the conclusions in three sections, each corresponding to a part of this dissertation.

7.1 Convex Relaxation of Unit Commitment and its Applications

We have proposed a polynomially-solvable primal formulation for the Lagrangian dual problem of the UC problem. This primal formulation explicitly describes the convex hull of each unit's feasible set and the convex envelope of each unit's cost function. We cast our formulation as a second-order cone program if the cost functions are quadratic, and as a linear program if the cost functions are affine or piecewise linear.

We then apply our primal formulation to convex hull pricing. We show that exact convex hull prices can be obtained in the absence of ramping constraints, and that exactness is preserved when we consider ancillary services or any linear system-wide constraints. A tractable approximation applies when ramping constraints are considered. A 96-period 76-unit transmission-

constrained example solves in less than fifteen seconds on a personal computer. This example shows that prices obtained through our formulation further reduce uplift payments compared to a single-period approximation of convex hull pricing.

Next, we apply our primal formulation to generation expansion planning (GEP). We propose a GEP model that incorporates commitment decisions in a tractable manner by employing our convex relaxation of UC as the operational model. We apply our GEP model to a Texas system in which we model hourly system operations in the year of 2030. We demonstrate that neglecting commitment decisions in GEP results in an under-investment in flexible generation, which may lead to reserve shortage, load shedding, and curtailment of renewable generation. We show that our convex relaxation outperforms an existing unit-clustering method in terms of accuracy. We believe that our approach might be even more advantageous in terms of accuracy when transmission constraints are considered, since clustering similar units that are not co-located introduces additional error in the presence of congestion.

Our convex relaxation could be used in future research for other power system optimization problems in which commitment decisions need to be represented. The convex envelopes presented in this dissertation may also be used to accelerate the solution of UC. Characterizing the convex hull of an individual unit's feasible set in the presence of additional complications (e.g., ramping constraints, startup types) is another research direction.

Our GEP model embeds a convex operational problem, lending itself

to Benders decomposition. This decomposition approach could be used in a transmission-constrained context and include additional features such as, for example, Monte Carlo simulation of outages.

7.2 Modeling of Combined Cycle Units

We have presented a mixed-integer programming formulation of combined-cycle units with explicit representation of transition ramping. This formulation removes a common assumption in the existing literature: the completion of any transition within a single interval. We show the tightness and compactness of the proposed model. We also present a computational study on a test system.

We have showed that if we have the characterization of the convex hull of the feasible binary transition decisions, then the characterization of the entire feasible set of an individual unit is trivial. However, the characterization of the convex hull of the binary decision needs to be explored in the future in the presence of minimum up/down time constraints.

Large-scale computational studies in the future can quantify the benefit of the proposed model in practice. The proposed configuration-based model can potentially apply to any unit with multiple operational modes, e.g., pumped-hydro storage plants.

7.3 Pricing in Multi-Interval Real-Time Markets

We have considered a multi-interval real-time market formulated as a look-ahead dispatch (LAD) problem implemented in a model predictive control fashion. If an independent system operator (ISO) adopts prices obtained from the Lagrange multipliers of this LAD problem, market participants may have an incentive to deviate from the ISO's dispatch because historical losses are treated as sunk. To mitigate incentives to deviate, we describe two methods that incorporate historical loss: an existing method that dualizes the system-wide constraints and a new method that dualizes inter-temporal constraints.

Both methods produce prices that support the ISO's optimal dispatch given a fixed finite horizon with perfect foresight. In more practical scenarios with a rolling horizon, we use two types of uplift payment as qualitative measures of incentive: make-whole payment and lost opportunity cost. Computational studies on an ISO New England-based system show that the proposed method results in lower out-of-market payments and a welfare transfer from consumers to suppliers.

Although our formulations allow general linear inter-temporal constraints, we have mainly considered ramping constraints in the computational studies. If an ISO were to manage the state-of-charge of energy storage resources, new inter-temporal equality constraints would need to be introduced to reflect the change of state-of-charge from one interval to the next. Pricing in the presence of inter-temporal operations of storage resources is one future research direction.

Appendices

Appendix A

A Proof of Lemma 2.9 in [69]

Over a time horizon $T \in \mathbb{Z}_{++}$, consider a generating unit with minimum up time of $L \in \mathbb{Z}_+$ dispatch intervals and minimum down time of $l \in \mathbb{Z}_+$ dispatch intervals. Let $\boldsymbol{x}$ (denoted by $\boldsymbol{u}$ in [69]) be the commitment vector and $\boldsymbol{u}$ (denoted by $\boldsymbol{v}$ in [69]) be the start-up vector.

The commitment polytope of a unit $D_T(L, l)$ is defined to be

$$
\begin{aligned}
D_T(L,l) = \{\boldsymbol{x} \in \mathbb{R}^T, \boldsymbol{u} \in \mathbb{R}^{T-1} \mid & \sum_{i=t-L+1}^{t} u_i \leq x_t, && \forall t \in [L+1, T], && \text{(A.1)}\\
& \sum_{i=t-l+1}^{t} u_i \leq 1 - x_{t-l}, && \forall t \in [l+1, T], && \text{(A.2)}\\
& u_t \geq x_t - x_{t-1}, && \forall t \in [2, T], && \text{(A.3)}\\
& u_t \geq 0, && \forall t \in [2, T]\}. && \text{(A.4)}
\end{aligned}
$$

To simplify our proof, let $\boldsymbol{w}$ be the shut-down vector, so that

$$w_t = u_t + x_{t-1} - x_t, \forall t \in [2, T]. \tag{A.5}$$

Lemma 2.9 in [69]:

Lemma A.1. *Let $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}) \in D_T(L, l)$. Then there exist integral points $a^s \in D_T(L, l), s \in \mathcal{S}$, and $\lambda_s \in \mathbb{R}_+, s \in \mathcal{S}$ such that*

(i). $\overline{\boldsymbol{x}} = \sum_{s\in\mathcal{S}} \lambda_s \boldsymbol{x}(a^s)$, $\overline{\boldsymbol{u}} = \sum_{s\in\mathcal{S}} \lambda_s \boldsymbol{u}(a^s)$, *and* $\sum_{s\in\mathcal{S}} \lambda_s = 1$;

(ii). *let* $\mathcal{S}_t^d$ *be the set of all points that have been shut down at* t*, then* $\forall t \in [2, T]$, $\overline{w}_t = \sum_{s\in\mathcal{S}_t^d} \lambda_s$;

(iii). *let* $\mathcal{S}_t^u$ *be the set of all points that have been started up at* t*, then* $\forall t \in [2, T]$, $\overline{u}_t = \sum_{s\in\mathcal{S}_t^u} \lambda_s$,

where $\boldsymbol{u}(a^s)$ *is the* $\boldsymbol{u}$ *vector corresponding to* a^s.

This lemma states that any point in $D_T(L, l)$ can be written as a convex combination of a set of integral points in this polytope. This implies that every extreme point of $D_T(L, l)$ is integral.

The proof of this lemma provided in [69] has the following flaws:

- the proof by induction uses a base case of $D_2(1, 1)$ and an inductive step that shows the statements are true for $D_T(L, l)$ if they are true for $D_{T-1}(L, l)$. Multiple natural numbers are varying, and induction must be applied to T, L and l;
- in the inductive step, the authors consider "$\overline{u}_t$ of the integral points", which is not well-defined, since a fraction of a natural number can be fractional;
- the determination of λ_s is not specified.

We state the lemma in a slightly different way and give a proof of the new lemma.

Lemma A.2. *For all* $T \in \mathbb{Z}_{++}\backslash\{1\}$, $\forall L, l \in \mathbb{Z}_+$ *such that* $L \leq T - 1$ *and*

$l \leq T-1$, $\forall (\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}) \in D_T(L,l)$, *there exists a set of integral points* $\{(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s), s \in \mathcal{S} = \{1,2,\ldots,|\mathcal{S}|\}\} \subset D_T(L,l)$, *and* $\{\lambda^s \in \mathbb{R}_+, s \in \mathcal{S}\}$ *such that*

(i). $\overline{\boldsymbol{x}} = \sum_{s\in\mathcal{S}} \lambda^s \hat{\boldsymbol{x}}^s$, $\overline{\boldsymbol{u}} = \sum_{s\in\mathcal{S}} \lambda^s \hat{\boldsymbol{u}}^s$, *and* $\sum_{s\in\mathcal{S}} \lambda^s = 1$;

(ii). *for all* $t \in [2,T]$, $\overline{w}_t = \sum_{s\in\mathcal{S}_{t,d}} \lambda^s$, *where* $\mathcal{S}_{t,d} = \{s \mid \hat{w}_t^s = 1\}$;

(iii). *for all* $t \in [2,T]$, $\overline{u}_t = \sum_{s\in\mathcal{S}_{t,u}} \lambda^s$, *where* $\mathcal{S}_{t,u} = \{s \mid \hat{u}_t^s = 1\}$.

Proof. We prove by induction.

Base case: consider $T = 2$. Since the time resolution we consider is one dispatch interval, a minimum up/down time of less than one dispatch interval has the same effect as a minimum up/down time of one dispatch interval. Therefore, $D_2(1,1) = D_2(1,0) = D_2(0,1) = D_2(0,0)$, so that we can only consider the case $D_2(1,1)$, which is the case considered in the original proof of [69].

Induction hypothesis: suppose the given statement holds for $T-1$, where $T > 2$ and $T \in \mathbb{Z}$. That is, suppose $\forall L', l' \in \mathbb{Z}_+$ such that $L' \leq T-2$ and $l' \leq T-2$, $\forall (\boldsymbol{x}, \boldsymbol{u}) \in D_{T-1}(L', l')$, there exists a set of integral points $\{(\tilde{\boldsymbol{x}}^s, \tilde{\boldsymbol{u}}^s), s \in \mathcal{S}' = \{1,2,\ldots,|\mathcal{S}'|\}\} \subset D_{T-1}(L', l')$, and $\{\mu^s \in \mathbb{R}_+, s \in \mathcal{S}'\}$ such that (i), (ii), and (iii) hold.

We need to show: the statement holds for T. That is, $\forall L, l \in \mathbb{Z}_+$ such that $L \leq T-1$ and $l \leq T-1$, $\forall (\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}) \in D_T(L,l)$, there exists a set of integral points $\{(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s), s \in \mathcal{S} = \{1,2,\ldots,|\mathcal{S}|\}\} \subset D_T(L,l)$, and $\{\lambda^s \in \mathbb{R}_+, s \in \mathcal{S}\}$ such that (i), (ii), and (iii) hold.

Given any $L, l \in \mathbb{Z}_{++}$ such that $L \leq T-1$ and $l \leq T-1$, and given any $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}) \in D_T(L, l)$, we drop the last entry of $\overline{\boldsymbol{x}}$ and denote the truncated vector by $\boldsymbol{x} \in \mathbb{R}^{T-1}$. Similarly, we drop the last entry of $\overline{\boldsymbol{u}}$ and denote the truncated vector as $\boldsymbol{u} \in \mathbb{R}^{T-2}$. We then have $(\boldsymbol{x}, \boldsymbol{u}) \in D_{T-1}(L-1, l-1)$.[1] Since $L-1 \leq T-2$ and $l-1 \leq T-2$, by the induction hypothesis, we can find a set of integral points $\{(\tilde{\boldsymbol{x}}^s, \tilde{\boldsymbol{u}}^s), s \in \mathcal{S}'\}$ and $\{\mu^s \in \mathbb{R}_+, s \in \mathcal{S}'\}$ that satisfy (i), (ii), and (iii). Similarly, we let $\mathcal{S}'_{t,d} = \{s \,|\, \tilde{w}^s_t = 1\}$ and $\mathcal{S}'_{t,u} = \{s \,|\, \tilde{u}^s_t = 1\}$.

We construct $\{(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s), s \in \mathcal{S}\} \subset D_T(L, l)$ from $\{(\tilde{\boldsymbol{x}}^s, \tilde{\boldsymbol{u}}^s), s \in \mathcal{S}'\} \subset D_{T-1}(L-1, l-1)$ by defining all but the last components of $\hat{\boldsymbol{x}}^s$ and $\hat{\boldsymbol{u}}^s$ to be the same as $\tilde{\boldsymbol{x}}^s$ and $\tilde{\boldsymbol{u}}^s$, respectively, and appending a T-th component as needed. When necessary, we may construct more than one $(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s)$ based on a single $(\tilde{\boldsymbol{x}}^s, \tilde{\boldsymbol{u}}^s)$. We construct $(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s)$ in a way that allows us to find a set of λ^s satisfying (i), (ii), and (iii).

To facilitate our proof, we partition $\mathcal{S}'$ into $\mathcal{S}'_1$ and $\mathcal{S}'_2$, so that

$$\mathcal{S}'_1 = \{s \in \mathcal{S}' \,|\, \tilde{x}^s_{T-1} = 1\}, \tag{A.6}$$

and

$$S'_2 = \{s \in \mathcal{S}' \,|\, \tilde{x}^s_{T-1} = 0\}. \tag{A.7}$$

That is, $\mathcal{S}'_1$ contains integral points that involve the unit being on at time $T-1$.

[1] For the boundary cases where either L or l equals one, or both, note that $D_T(L, 0) = D_T(L, 1)$, $D_T(0, l) = D_T(1, l)$, and $D_T(0, 0) = D_T(1, 1)$.

We further partition $\mathcal{S}'_1$ into $\mathcal{S}'_{11}$ and $\mathcal{S}'_{12}$, so that

$$\mathcal{S}'_{11} = \{s \in \mathcal{S}'_1 \,|\, \tilde{u}^s_{T-L+1} = \tilde{u}^s_{T-L+2} = \cdots = \tilde{u}^s_{T-1} = 0\}. \tag{A.8}$$

and $\mathcal{S}'_{12} = \mathcal{S}'_1 \backslash \mathcal{S}'_{11}$. That is, the integral points in $\mathcal{S}'_{11}$ involve the unit not starting up during $t \in [T - L + 1, T - 1]$.

Similarly, we partition $\mathcal{S}'_2$ into $\mathcal{S}'_{21}$ and $\mathcal{S}'_{22}$, so that

$$\mathcal{S}'_{21} = \{s \in \mathcal{S}'_1 \,|\, \tilde{w}^s_{T-l+1} = \tilde{w}^s_{T-l+2} = \cdots = \tilde{w}^s_{T-1} = 0\}, \tag{A.9}$$

and $\mathcal{S}'_{22} = \mathcal{S}'_2 \backslash \mathcal{S}'_{21}$. That is, the integral points in $\mathcal{S}'_{21}$ involve the unit not shutting down during $t \in [T - l + 1, T - 1]$.

We have effectively partitioned $\mathcal{S}'$ into $\mathcal{S}'_{11}$, $\mathcal{S}'_{12}$, $\mathcal{S}'_{21}$, and $\mathcal{S}'_{22}$, which results in the following properties:

- $\sum_{s\in\mathcal{S}'_1} \mu^s = x_{T-1}$, and $\sum_{s\in\mathcal{S}'_2} \mu^s = 1 - x_{T-1}$.
- $\sum_{s\in\mathcal{S}'_{11}} \mu^s \geq \overline{w}_T$.

  To see this, notice that $\sum_{s\in\mathcal{S}'_{11}} \mu^s = x_{T-1} - \sum_{t=T-L+1}^{T-1} \sum_{s\in\mathcal{S}'_{t,u}} \mu^s = x_{T-1} - \sum_{t=T-L+1}^{T-1} u_t$, where the first equality follows the definition of $\mathcal{S}'_{11}$ and $\mathcal{S}'_{t,u}$, and the second equality follows from (iii). Now we invoke the turn on inequality at T for $D_T(L, l)$: $\sum_{t=T-L+1}^{T} \overline{u}_t = \sum_{t=T-L+1}^{T-1} u_t + \overline{u}_T \leq \overline{x}_T$. Applying this turn on inequality to the previous equality yields $\sum_{s\in\mathcal{S}'_{11}} \mu^s \geq -\overline{x}_T + \overline{x}_{T-1} + \overline{u}_T$. The desired result follows from (A.5).
- $\sum_{s\in\mathcal{S}'_{21}} \mu^s \geq \overline{u}_T$.

To see this, notice that $\sum_{s\in\mathcal{S}'_{21}}\mu^s = 1-x_{T-1}-\sum_{t=T-l+1}^{T-1}\sum_{s\in\mathcal{S}'_{t,d}}\mu^s = 1-x_{T-1}-\sum_{t=T-l+1}^{T-1}w_t$, where the first equality is by definition of $\mathcal{S}'_{21}$ and $\mathcal{S}'_{t,d}$, and the second equality follows from (ii). Now we invoke the turn off inequality at T for $D_T(L,l)$: $\sum_{t=T-l+1}^{T}\overline{w}_t = \sum_{t=T-l+1}^{T-1}w_t+\overline{w}_T \le 1-\overline{x}_T$. Applying this turn off inequality to the previous equality yields $\sum_{s\in\mathcal{S}'_{21}}\mu^s \ge \overline{x}_T-\overline{x}_{T-1}+\overline{w}_T$. The desired result follows from (A.5).

To satisfy (ii), we would like to append a one to some of the shut-down vectors $\tilde{\boldsymbol{w}}^s$ and assign positive values to their associated λ^s, so that $\overline{w}_T = \sum_{s\in\mathcal{S}_{T,d}}\lambda^s$. Because of the minimum up time constraints, we can only append a one to those $\tilde{\boldsymbol{w}}^s$ with $s\in\mathcal{S}'_{11}$.

To satisfy (iii), we would like to append a one to some of the start-up vectors $\tilde{\boldsymbol{u}}^s$ and assign positive values to their associated λ^s, so that $\overline{u}_T = \sum_{s\in\mathcal{S}_{T,u}}\lambda^s$. Because of the minimum down time constraints, we can only append a one to those $\tilde{\boldsymbol{u}}^s$ with $s\in\mathcal{S}'_{21}$.

We construct $\{(\hat{\boldsymbol{x}}^s,\hat{\boldsymbol{u}}^s), s\in\mathcal{S}\}$ according to the following rules:

- Construct two integral points $(\hat{\boldsymbol{x}}^{s1},\hat{\boldsymbol{u}}^{s1}), (\hat{\boldsymbol{x}}^{s2},\hat{\boldsymbol{u}}^{s2})$ from each $\{(\tilde{\boldsymbol{x}}^s,\tilde{\boldsymbol{u}}^s), s\in\mathcal{S}'_{11}\}$. Each shut-down vector $\hat{\boldsymbol{w}}^{s1}$ is created by appending a one to $\tilde{\boldsymbol{w}}^s$, and each $\hat{\boldsymbol{w}}^{s2}$ is created by appending a zero. Each start-up vector $\hat{\boldsymbol{u}}^{s1}$ and $\hat{\boldsymbol{u}}^{s2}$ can only be created by appending a zero to $\tilde{\boldsymbol{u}}^s$. Each commitment vector $\hat{\boldsymbol{x}}^{s1}$ is created by appending a zero to $\tilde{\boldsymbol{x}}^s$, and each $\hat{\boldsymbol{x}}^{s2}$ is created by appending a one. We set

$\lambda^{s1} + \lambda^{s2} = \mu^s, \forall s \in \mathcal{S}'_{11}$. Because $\sum_{s\in\mathcal{S}'_{11}} \mu^s \geq \overline{w}_T$, we can find $\lambda^{s1}, \lambda^{s2}, s \in \mathcal{S}'_{11}$ such that $\sum_{s\in\mathcal{S}'_{11}} \lambda^{s1} = \overline{w}_T$.

- Construct only one integral point $(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s)$ from each $\{(\tilde{\boldsymbol{x}}^s, \tilde{\boldsymbol{u}}^s), s \in \mathcal{S}'_{12}\}$. Because of the minimum up time constraints, we can only append a one to $\tilde{\boldsymbol{x}}^s$, a zero to $\tilde{\boldsymbol{u}}^s$, and a zero to $\tilde{\boldsymbol{w}}^s$. We set $\lambda^s = \mu^s, \forall s \in \mathcal{S}'_{12}$.
- Construct two integral points $(\hat{\boldsymbol{x}}^{s1}, \hat{\boldsymbol{u}}^{s1}), (\hat{\boldsymbol{x}}^{s2}, \hat{\boldsymbol{u}}^{s2})$ from each $\{(\tilde{\boldsymbol{x}}^s, \tilde{\boldsymbol{u}}^s), s \in \mathcal{S}'_{21}\}$. We create each start-up vector $\hat{\boldsymbol{u}}^{s1}$ by appending a one to $\tilde{\boldsymbol{u}}^s$, and each $\hat{\boldsymbol{u}}^{s2}$ by appending a zero. Each shut-down vector $\hat{\boldsymbol{w}}^{s1}$, and $\hat{\boldsymbol{w}}^{s2}$ can only be created by appending a zero to $\tilde{\boldsymbol{w}}^s$. Each commitment vector $\hat{\boldsymbol{x}}^{s1}$ is created by appending a one to $\tilde{\boldsymbol{x}}^s$, and $\hat{\boldsymbol{x}}^{s2}$ is created by appending a zero. We set $\lambda^{s1} + \lambda^{s2} = \mu^s, \forall s \in \mathcal{S}'_{21}$. Because $\sum_{s\in\mathcal{S}'_{21}} \mu^s \geq \overline{u}_T$, we can find $\lambda^{s1}, \lambda^{s2}, s \in \mathcal{S}'_{21}$ such that $\sum_{s\in\mathcal{S}'_{21}} \lambda^{s1} = \overline{u}_T$.
- Construct only one integral point $(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s)$ from each $\{(\tilde{\boldsymbol{x}}^s, \tilde{\boldsymbol{u}}^s), s \in \mathcal{S}'_{22}\}$. Because of the minimum down time constraints, we can only append a zero to $\tilde{\boldsymbol{x}}^s$, a zero to $\tilde{\boldsymbol{u}}^s$, and a zero to $\tilde{\boldsymbol{w}}^s$. We set $\lambda^s = \mu^s, \forall s \in \mathcal{S}'_{22}$.

To summarize, we have constructed $\{(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s), s \in \mathcal{S}\} = \{(\hat{\boldsymbol{x}}^{s1}, \hat{\boldsymbol{u}}^{s1}), s \in \mathcal{S}'_{11}\} \cup (\hat{\boldsymbol{x}}^{s2}, \hat{\boldsymbol{u}}^{s2}), s \in \mathcal{S}'_{11}\} \cup \{(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s), s \in \mathcal{S}'_{12}\} \cup \{(\hat{\boldsymbol{x}}^{s1}, \hat{\boldsymbol{u}}^{s1}), s \in \mathcal{S}'_{21}\} \cup \{(\hat{\boldsymbol{x}}^{s2}, \hat{\boldsymbol{u}}^{s2}), s \in \mathcal{S}'_{21}\} \cup \{(\hat{\boldsymbol{x}}^s, \hat{\boldsymbol{u}}^s), s \in \mathcal{S}'_{22}\}$ and their associated λ.

Finally, we verify (i), (ii), (iii) using the integral points and $\lambda^s, s \in \mathcal{S}$ that we have constructed. Since we keep the first $T-1$ components of each a^s

to be the same as its corresponding p^s, by the way we construct λ^s, it suffices to verify (i), (ii), (iii) for only $t = T$.

(i). We have $\sum_{s\in\mathcal{S}} \lambda^s = 1$. It suffices to show that $\overline{x}_t = \sum_{s\in\mathcal{S}} \lambda^s \hat{x}^s_T$, $\overline{u}_T = \sum_{s\in\mathcal{S}} \lambda^s \hat{u}^s_T$.

By construction, we have $\sum_{s\in\mathcal{S}} \lambda^s \hat{x}^s_T = \sum_{s\in\mathcal{S}'_{11}} \lambda^{s2} + \sum_{s\in\mathcal{S}'_{12}} \lambda^s + \sum_{s\in\mathcal{S}'_{21}} \lambda^{s1} = x_{T-1} - \overline{w}_T + \overline{u}_T = \overline{x}_t$.

Also, we have $\sum_{s\in\mathcal{S}} \lambda^s \hat{u}^s_T = \sum_{s\in\mathcal{S}'_{21}} \lambda^{s1} = \overline{u}_T$.

(ii). By construction, we have $\sum_{s\in\mathcal{S}_{T,d}} \lambda^s = \sum_{s\in\mathcal{S}'_{11}} \lambda^{s1} = \overline{w}_T$.

(iii). By construction, we have $\sum_{s\in\mathcal{S}_{T,u}} \lambda^s = \sum_{s\in\mathcal{S}'_{21}} \lambda^{s1} = \overline{u}_T$.

□

Bibliography

[1] Yousef M. Al-Abdullah, Ahmed Salloum, Kory W. Hedman, and Vijay Vittal. Analyzing the Impacts of Constraint Relaxation Practices in Electric Energy Markets. *IEEE Transactions on Power Systems*, 31(4):2566–2577, July 2016.

[2] Sami Ammari and Kwok W. Cheung. Advanced combined-cycle modeling. *2013 IEEE Grenoble Conference PowerTech, POWERTECH 2013*, 2013.

[3] George Anders and Atef Morched. Commitment techniques for combined-cycle generating units. Technical report, 2005.

[4] Egon Balas. Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Applied Mathematics*, 89(1-3):3–44, 1998.

[5] Jonathan F. Bard. Short-Term Scheduling of Thermal-Electric Generators Using Lagrangian Relaxation. *Operations Research*, 36(5):756–766, October 1988.

[6] Carlos Batlle and Pablo Rodilla. An Enhanced Screening Curves Method for Considering Thermal Cycling Operation Costs in Generation Expansion Planning. *IEEE Transactions on Power Systems*, 28(4):3683–3691, November 2013.

[7] Dimitri Bertsekas and Nils Sandell. Estimates of the duality gap for large-scale separable nonconvex optimization problems. In *1982 21st IEEE Conference on Decision and Control*, pages 782–785, 1982.

[8] Dimitri P. Bertsekas, Gregory S. Lauer, Nils R. Sandell, and Thomas A. Posbergh. Optimal Short-Term Scheduling of Large-Scale Power Systems. *IEEE Transactions on Automatic Control*, 28(1):1–11, 1983.

[9] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[10] Michael Cadwalader, Paul Gribik, William Hogan, and Susan Pope. Extended LMP and Financial Transmission Rights. Technical report, Harvard University, 2010.

[11] M. Carrion and J.M. Arroyo. A Computationally Efficient Mixed-Integer Linear Formulation for the Thermal Unit Commitment Problem. *IEEE Transactions on Power Systems*, 21(3):1371–1378, August 2006.

[12] G. W. Chang, G. S. Chuang, and T. K. Lu. A simplified combined-cycle unit model for mixed integer linear programming-based unit commitment. In *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–6. IEEE, July 2008.

[13] Dae-Hyun Choi and Le Xie. Data Perturbation-Based Sensitivity Analysis of Real-Time Look-Ahead Economic Dispatch. *IEEE Transactions*

on Power Systems, 32(3):2072–2082, May 2017.

[14] M Conforti, G Cornuejols, and G Zambelli. *Integer Programming*. Graduate Texts in Mathematics. Springer, 2014.

[15] Cong Liu, M. Shahidehpour, Zuyi Li, and M. Fotuhi-Firuzabad. Component and Mode Models for the Short-Term Scheduling of Combined-Cycle Units. *IEEE Transactions on Power Systems*, 24(2):976–990, May 2009.

[16] Gérard Cornuéjols. Valid inequalities for mixed integer linear programs. *Mathematical Programming*, 112(1):3–44, 2007.

[17] Chenxi Dai, Yonghong Chen, Fengyu Wang, Jie Wan, and Lei Wu. A Tight Configuration-Component Based Hybrid Model for Combined-Cycle Units in MISO Day-Ahead Market. August 2017.

[18] Pelin Damcı-Kurt, Simge Küçükyavuz, Deepak Rajan, and Alper Atamtürk. A polyhedral study of production ramping. *Mathematical Programming*, 158(1-2):175–205, July 2016.

[19] Cedric De Jonghe, Benjamin F. Hobbs, and Ronnie Belmans. Optimal Generation Mix With Short-Term Demand Response and Wind Penetration. *IEEE Transactions on Power Systems*, 27(2):830–839, 2012.

[20] Erik Ela and Mark O'Malley. Scheduling and Pricing for Expected Ramp Capability in Real-Time Power Markets. *IEEE Transactions on Power Systems*, 31(3):1681–1691, May 2016.

[21] EPIS. AURORAxmp Database, Accessed on April 09, 2016.

[22] ERCOT. ERCOT Protocols Section 4: Day-Ahead Operations, 2016.

[23] James E. Falk. Lagrange Multipliers and Nonconvex Programs. *SIAM Journal on Control*, 7(4):534–545, 1969.

[24] Lei Fan and Yongpei Guan. An Edge-Based Formulation for Combined-Cycle Units. *IEEE Transactions on Power Systems*, 31(3):1809–1819, May 2016.

[25] Federal Energy Regulatory Commission. Price Formation in Energy and Ancillary Services Markets Operated by Regional Transmission Organizations and Independent System Operators, 2015.

[26] FERC. Price Formation in Energy and Ancillary Services Markets Operated by Regional Transmission Organizations and Independent System Operators. Technical report, FERC, 2015.

[27] Jean-Louis Goffin and Jean-Philippe Vial. Convex nondifferentiable optimization: A survey focused on the analytic center cutting plane method. *Optimization Methods and Software*, 17(5):805–867, 2002.

[28] Michael Grant and Stephen Boyd. CVX: Matlab Software for Disciplined Convex Programming, version 2.1, 2014.

[29] Paul Gribik, William Hogan, and Susan Pope. Market-Clearing Electricity Prices and Energy Uplift. Technical report, Harvard University, 2007.

[30] X. Guan, P.B. Luh, H. Yan, and J.A. Amalfi. An optimization-based method for unit commitment. *International Journal of Electrical Power & Energy Systems*, 14(1):9–17, February 1992.

[31] Gurobi Optimization Inc. Gurobi optimizer reference manual, 2016.

[32] X.S. Han, H.B. Gooi, and D.S. Kirschen. Dynamic economic dispatch: feasible and optimal solutions. *IEEE Transactions on Power Systems*, 16(1):22–28, 2001.

[33] William Hogan. Electricity Market Design and Efficient Pricing: Applications for New England and Beyond. Technical report, Harvard University, 2014.

[34] William Hogan and Brendan Ring. On Minimum-Uplift Pricing for Electricity Markets. Technical report, Harvard University, 2003.

[35] William W. Hogan. Electricity Market Design: Optimization and Market Equilibrium, 2016.

[36] B. Hua, R. Baldick, and J. Wang. Representing Operational Flexibility in Generation Expansion Planning Through Convex Relaxation of Unit Commitment. *IEEE Transactions on Power Systems*, 33(2), 2018.

[37] Bowen Hua and Ross Baldick. A convex primal formulation for convex hull pricing. *IEEE Transactions on Power Systems*, 32(5):3814–3823, 2017.

[38] Hailong Hui, Chien-Ning Yu, Feng Gao, and Resmi Surendran. Combined cycle resource scheduling in ERCOT nodal market. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–8. IEEE, July 2011.

[39] R. J. Kaye and H. R. Outhred. A theory of electricity tariff design for optimal operation and investment. *IEEE Transactions on Power Systems*, 4(2):606–613, May 1989.

[40] Dheepak Krishnamurthy, Wanning Li, and Leigh Tesfatsion. An 8-Zone Test System Based on ISO New England Data: Development and Application. *IEEE Transactions on Power Systems*, 31(1):234–246, 2016.

[41] Eamonn Lannoye, Damian Flynn, and Mark O'Malley. The role of power system flexibility in generation planning. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–6. IEEE, July 2011.

[42] C. Lemaréchal and A. Renaud. A geometric study of duality gaps, with applications. *Mathematical Programming*, 90(3):399–427, 2001.

[43] Claude Lemaréchal. Lagrangian relaxation. In *Computational combinatorial optimization*, pages 112–156. Springer, 2001.

[44] Tao Li and Mohammad Shahidehpour. Price-based unit commitment: A case of Lagrangian relaxation versus mixed integer programming. *IEEE Transactions on Power Systems*, 20(4):2015–2025, 2005.

[45] George Liberopoulos and Panagiotis Andrianesis. Critical Review of Pricing Schemes in Markets with Non-Convex Costs. *Operations Research*, 64(1):17–31, 2016.

[46] Haifeng Liu, Leigh Tesfatsion, and A. A. Chowdhury. Locational marginal pricing basics for restructured wholesale power markets. *IEEE Power and Energy Society General Meeting*, 2009.

[47] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1-3):193–228, 1998.

[48] Juan Ma, Vera Silva, Regine Belhomme, Daniel S. Kirschen, and Luis F. Ochoa. Evaluating and planning flexibility in sustainable power systems. In *IEEE Transactions on Sustainable Energy*, volume 4, pages 200–209. IEEE, 2013.

[49] Andreu Mas-Colell, Michael Dennis Whinston, and Jerry R Green. *Microeconomic Theory*. Oxford University Press, 1995.

[50] Joel Mickey. Multi-Interval Real-Time Market Overview, 2015.

[51] MISO. Convex Hull Workshop 3. Technical report, 2010.

[52] German Morales-España, Carlos M. Correa-Posada, and Andres Ramos. Tight and Compact MIP Formulation of Configuration-Based Combined-Cycle Units. *IEEE Transactions on Power Systems*, 31(2):1350–1359, March 2016.

[53] German Morales-España, Jesus M. Latorre, and Andres Ramos. Tight and Compact MILP Formulation of Start-Up and Shut-Down Ramping in Unit Commitment. *IEEE Transactions on Power Systems*, 28(2):1288–1296, May 2013.

[54] Germán Morales-España, Jesus M. Latorre, and Andres Ramos. Tight and compact MILP formulation for the thermal unit commitment problem. *IEEE Transactions on Power Systems*, 28(4):4897–4908, 2013.

[55] John A. Muckstadt and Sherri A. Koenig. An Application of Lagrangian Relaxation to Scheduling in Power-Generation Systems. *Operations Research*, 25(3):387–403, June 1977.

[56] C. I. Nweke, F. Leanez, G. R. Drayton, and M. Kolhe. Benefits of chronological optimization in capacity planning for electricity markets. In *2012 IEEE International Conference on Power System Technology (POWERCON)*, pages 1–6. IEEE, October 2012.

[57] Richard P O'Neill, Paul M Sotkiewicz, Benjamin F Hobbs, Michael H Rothkopf, and William R Stewart. Efficient market-clearing prices in markets with nonconvexities. *European Journal of Operational Research*, 164(1):269–285, 2005.

[58] André Ortner and Daniel Huppmann. Modeling competitive equilibrium prices for energy and balancing capacity in electricity markets involving non-convexities, 2016.

[59] James Ostrowski, Miguel F. Anjos, and Anthony Vannelli. Tight Mixed Integer Linear Programming Formulations for the Unit Commitment Problem. *IEEE Transactions on Power Systems*, 27(1):39–46, 2012.

[60] Bryan Palmintier and Mort Webster. Impact of unit commitment constraints on generation expansion planning with renewables. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–7. IEEE, 2011.

[61] Bryan S. Palmintier. *Incorporating Operational Flexibility into Electric Generation Planning - Impacts and Methods for System Design and Policy Analysis*. PhD thesis, Massachusetts Institute of Technology, 2013.

[62] Bryan S. Palmintier and Mort D. Webster. Heterogeneous unit clustering for efficient operational flexibility modeling. *IEEE Transactions on Power Systems*, 29(3):1089–1098, 2014.

[63] Bryan S. Palmintier and Mort D. Webster. Impact of Operational Flexibility on Electricity Generation Planning With Renewable and Carbon Targets. *IEEE Transactions on Sustainable Energy*, 7(2):672–684, 2016.

[64] Kai Pan and Yongpei Guan. A Polyhedral Study of the Integrated Minimum-Up/-Down Time and Ramping Polytope. Technical report, University of Florida, 2015.

[65] Tengshun Peng and Dhiman Chatterjee. Pricing Mechanism for Time Coupled Multi-interval Real-Time Dispatch. In *FERC Technical Conference*, 2013.

[66] D Phillips and Fp Jenkin. A mathematical model for determining generating plant mix. *Third PSCC, Rome*, 1969.

[67] PJM. PJM Manual 11: Energy and Ancillary Services Market Operations, 2015.

[68] James E. Price and Mark Rothleder. Recognition of extended dispatch horizons in California's energy markets. In *2011 IEEE Power and Energy Society General Meeting*, pages 1–5. IEEE, July 2011.

[69] Deepak Rajan and Samer Takriti. Minimum up/down polytopes of the unit commitment problem with start-up costs. Technical report, IBM Research Division, 2005.

[70] Brendan Ring. *Dispatch based pricing in decentralised power systems*. PhD thesis, University of Canterbury, 1995.

[71] Royal Academy of Engineering. The cost of generating electricity, 2004.

[72] Carlos Ruiz, Antonio J. Conejo, and Steven A. Gabriel. Pricing non-convexities in an electricity pool. *IEEE Transactions on Power Systems*, 27(3):1334–1342, 2012.

[73] Wolf Peter Schill, Michael Pahle, and Christian Gambardella. Start-up costs of thermal power plants in markets with increasing shares of variable renewable generation. *Nature Energy*, 2(6), 2017.

[74] Dane A. Schiro. Flexibility Procurement and Reimbursement: A Multi-period Pricing Approach. In *FERC Technical Conference*, 2017.

[75] Dane A. Schiro, Tongxin Zheng, Feng Zhao, and Eugene Litvinov. Convex Hull Pricing in Electricity Markets: Formulation, Analysis, and Implementation Challenges. *IEEE Transactions on Power Systems*, 31(5):4068 – 4075, 2016.

[76] Aonghus Shortt, Juha Kiviluoma, and Mark O'Malley. Accommodating Variability in Generation Planning. *IEEE Transactions on Power Systems*, 28(1):158–169, 2013.

[77] Fereidoom P. Sioshansi. *Competitive Electricity Markets Design, Implementation, Performance*. Else, 2011.

[78] Ramteen Sioshansi, Shmuel Oren, and Richard O'Neill. Three-part auctions versus self-commitment in day-ahead electricity markets. *Utilities Policy*, 18(4):165–173, 2010.

[79] Steven Stoft. *Power system economics*. 2002.

[80] Anupam A. Thatte, Dae-Hyun Choi, and Le Xie. Analysis of locational marginal prices in look-ahead economic dispatch. In *2014 Power Systems Computation Conference*, pages 1–7. IEEE, August 2014.

[81] US Energy Information Administration. Annual Energy Outlook 2015, 2015.

[82] Congcong Wang, Peter B. Luh, Paul Gribik, Tengshun Peng, and Li Zhang. Commitment Cost Allocation of Fast-Start Units for Approximate Extended Locational Marginal Prices. *IEEE Transactions on Power Systems*, 2016.

[83] Congcong Wang, Tengshun Peng, Peter B. Luh, Paul Gribik, and Li Zhang. The subgradient simplex cutting plane method for extended locational marginal prices. *IEEE Transactions on Power Systems*, 28(3):2758–2767, 2013.

[84] Gui Wang, Uday V. Shanbhag, Tongxin Zheng, Eugene Litvinov, and Sean Meyn. An Extreme-Point Subdifferential Method for Convex Hull Pricing in Energy and Reserve Markets—Part I: Algorithm Structure. *IEEE Transactions on Power Systems*, 28(3):2111–2120, 2013.

[85] Gui Wang, Uday V. Shanbhag, Tongxin Zheng, Eugene Litvinov, and Sean Meyn. An Extreme-Point Subdifferential Method for Convex Hull Pricing in Energy and Reserve Markets—Part II: Convergence Analysis and Numerical Performance. *IEEE Transactions on Power Systems*, 28(3):2121–2127, 2013.

[86] Laurence A Wolsey. *Integer programming*. Wiley, New York, 1998.

[87] Allen J. Wood and Bruce F. Wollenberg. *Power generation, operation, and control*. John Wiley & Sons, 3rd edition, 2012.

[88] Le Xie and Marija D. Ilic. Model predictive economic/environmental dispatch of power systems with intermittent resources. In *2009 IEEE Power & Energy Society General Meeting*. IEEE.

[89] Le Xie and Marija D. Ilic. Model predictive dispatch in electric energy systems with intermittent resources. In *2008 IEEE International Conference on Systems, Man and Cybernetics*, pages 42–47. IEEE, October 2008.

[90] X. Xu and R. Howard. Ramp rate modeling for ERCOT look ahead SCED. In *2013 IEEE Power & Energy Society General Meeting*, pages 1–5. IEEE, 2013.

[91] Tong Zhang, Ross Baldick, and Thomas Deetjen. Optimized generation capacity expansion using a further improved screening curve method. *Electric Power Systems Research*, 124:47–54, 2015.